We are extremely grateful for the helpful comments of the referees, which we have used to improve the manuscript as described below.

Reply to Anonymous Referee

1) Are the footprints calculated on a grid, and the gridded footprints used for the inversion, or are they averaged for a whole basin, and the basin footprint used for the inversions? The authors comment on sub-basin spatial pattern differences between EDGAR and VULCAN later in the paper, suggesting the latter, but it isn't stated anywhere that I can find. If it is, please point me to the right place.

Footprints were calculated on a grid (the resolution of gridded footprints are discussed in Section 2.3). When coupled with gridded emissions the resulting concentration is summed over each air basin, and emissions are scaled by air basin in the inversion. We have included more specific language in Section 2.3.

2) The temporal patterns in the prior uncertainty (i.e. are there any, or do the experiments assume a constant prior uncertainty for the entire year), as well as the spatial patterns in the prior uncertainty (is it fixed for the whole basin, or is it gridded within the basin?).

Prior uncertainty is specified relative to prior emissions, hence it differs in absolute magnitude for monthly differences in emissions. Over the state this variation is ~15% when comparing May/Oct-Nov to Jan-Feb (see Fig. 1d). Prior uncertainty is specified for the whole air basin. This is now clarified in Sections 2.2 and 2.5.3.

One assumption that is critical to the applicability of this paper to other studies with real data is an assessment of the interaction between the different types of uncertainties. In particular, spatial and temporal uncertainties in the prior are not independent from each other due to the effects of transport (though the signals here are diminished by the shorter run times). This is even more important for the experiments involving transport error, as the differences in footprints will interact with the differences in the temporal and spatial patterns. A limited set of experiments should be able to address whether these effects are significant, prior to having to do the full suite for multiple more cases.

The purpose of this paper was to quantify individual contributions that could potentially bias regional fossil fuel CO2 inversions. The reviewer makes an interesting suggestion to include experiments where the different sources of uncertainty are included to see how they interact. When we tried this the different sources of uncertainty interacted in various ways that were difficult to interpret and did not enable us to make general conclusions. The interactions can improve or worsen the results in different cases. Therefore, we prefer not to include these results in the paper. However, we have added to the discussion some comments on the individual nature of our experiments and how, in reality, multiple sources of uncertainty will interact.

"We note that while we have assessed individual contributions to uncertainty in the experiments formulated here, these contributions can also interact with each other. These interactions could

act to increase the resulting biases, or partly cancel them, depending on the combination used. The possibility for interacting effects implies that multiple prior emissions estimates and transport models should be used in inversions of real data."

Specific Comments:
What is the period of simulation?
The period of simulation is three 1 month long campaigns in May 2014, October-November 2014 and January-February 2015, analogous to the observations presented in Graven et al. 2018. We have added some clarification in lines 117-121. The transport model simulations were run 7 days back in time from time of observation (see lines 189, 197, 220-221).

211-213: The scaling of 0.5 is definitely a free parameter that deserves some sensitivity analysis as it could strongly affect your results. You could look at the covariance between the three transport models as a first guess. How does this covariance compare to the factor of 0.5*mean signal that you assume?
We use the factor of 0.5 to recreate the previous inversion setup of, for example, Fischer et al., 2017, Graven et al., 2018, to examine the posterior bias under these assumed initial conditions. We have added text to the methods and two results sections on the effect of varying the free parameter (or as I call it the uncertainty parameter) to 0.3, 0.5 and 0.8. The differences are only a few percent and they do not change our conclusions.

229-231: Correlation in the transport uncertainties assumed in your experimental setup should at least be examined before assuming your uncertainty covariance is diagonal. Again, you could use the 3-model ensemble to test this assumption.
Here we follow the commonly used practice of assuming the uncertainty covariance is diagonal (e.g. Gerbig et al. 2003; Zhao et al. 2009; Göckede et al. 2010; Jeong et al. 2012a; 2012b; 2013; Fischer et al., 2017; Graven et al., 2018). Although interesting, to test for non-diagonal covariance was deemed to be out of the scope of this study. However, we note that an experiment on the use of correlation in the model-measurement mismatch uncertainty matrix was carried out in Jeong et al., 2016. The study compared statewide posterior emissions of CH_4 in California estimated using a full (i.e. diagonal and non-diagonal) uncertainty matrix to a diagonal uncertainty matrix, and found that mean statewide emissions differed by ~6%. However this difference was not significant relative to the uncertainty (see Text 3 from the Supplemental of Jeong et al., 2016).

I have included in lines 591-596 the following:
"Here, we have assumed the model-measurement mismatch uncertainty matrix is diagonal, following previous work (e.g. Gerbig et al. 2003; Fischer et al., 2017), however the consideration of correlated errors in the uncertainty matrix has also been found to affect posterior emissions for methane in California and reduce their uncertainty at the level of several percent (Jeong et al. 2016)."

Additional references:

Zhao, C., A. E. Andrews, L. Bianco, J. Eluszkiewicz, A. Hirsch, C. MacDonald, T. Nehrkorn, and M. L. Fischer (2009), Atmospheric inverse estimates of methane emissions from Central California, J. Geophys. Res., 114, D16302, doi:10.1029/2008JD011671.

Jeong, S., C. Zhao, A. E. Andrews, L. Bianco, J. M. Wilczak, and M. L. Fischer (2012a), Seasonal variation of CH4 emissions from central California, J. Geophys. Res., 117, D11306, doi:10.1029/2011JD016896.

Jeong, S., C. Zhao, A. E. Andrews, E. J. Dlugokencky, C. Sweeney, L. Bianco, J. M. Wilczak, and M. L. Fischer (2012b), Seasonal variations in N2O emissions from central California. Geophys. Res. Lett., 39, L16805, doi:10.1029/2012GL052307.

Jeong, S., Newman, S., Zhang, J., Andrews, A. E., Bianco, L., Bagley, J., ... & LaFranchi, B. W. (2016). Estimating methane emissions in California's urban and rural regions using multitower observations. Journal of Geophysical Research: Atmospheres, 121(21).

Section 2.5.4: what is the prior flux estimate in this experiment?
Annually averaged Vulcan, see lines 334-335.

390 - 392: isn't this easy to test by looking at footprints?
Removed these lines.

393-401: I'm not sure I see the connection here. Why does being more concentrated in urban regions change the total? (this also applies to the conclusion in lines 530-533) It also appears that there might be a temporal offset happening, where the fluxes are biased low in summer and high in fall/winter. Is there a pattern in the prior uncertainty causing this? There isn't any way to tell given the lack of temporal information in the flux results images. Another cause is the seasonality in sensitivity of the observations to the fluxes, which can again be tested by looking at footprints.
It is due to the fact that EDGAR places more emissions in built up areas. Therefore simulating concentrations at observation sites in built up areas using EDGAR emissions will, on average, have higher simulated concentrations compared to Vulcan, while total emissions are the same over the whole air basin. This leads to a scaling down of the fluxes when observation sites are primarily in built-up areas when EDGAR is used as the prior. For example, the fluxes are consistently biased low in the South Coast (14.SC) in Fig 4b.

There is no temporal offset as emissions are temporally constant in all spatial error experiments. Likewise the prior uncertainty stays constant throughout spatial error experiments.

3.2.4: A nice conclusion of this section is that the removal of outliers improves the results from transport errors alone. That could be a strong recommendation to the community for working at these scales, which is done by many modelers, but not all. - this is mentioned in the discussion, but could be more strongly highlighted here.
Added highlight - thank you.

I would suggest that a more direct analysis of the impact of transport errors by season could be accomplished by looking at basin-wide sensitivity for each observation location by season, and how that varies by transport model. This would explain a lot of the inter-model differences you are seeing in many of the other experiments as well. That would support your PBLH analysis, which gets to the heart of why the footprints would be different, but doesn't quantify the differences between the flux sensitivities directly.

We show in Fig 2 the air basin sensitivity for each observation location by season and how that varies by transport model.

537-540: This conclusion needs to be tested by altering the estimate of transport error assumed in the inversions themselves. My guess is that the answers might be sensitive to this parameter, but that needs to be tested.

See previous answer for testing the free parameter for transport error.

Fig 2: What is "signal"? Is it just the emissions run forward through the transport

Yes - see definition in brackets on lines 225-228.

Reply to Sourish Basu

1. Line 58, "INFLUX ref" is missing.
Corrected.

2. Lines 64-65, the phrasing "simulated and observed measurements" sounds awkward to me. I understand the authors phrased it this way because the Fischer et al (2017) study is an OSSE study. I suggest rephrasing this as "Recent studies with both real atmospheric measurements of $\Delta_{14}CO_2$ and $\Delta_{14}CO_2$ simulated in observing system simulation experiments (OSSEs) at a network of sites have shown that atmospheric $\Delta_{14}CO_2$ can be used to estimate monthly mean Californian $ffCO_2$ emissions with posterior uncertainties of 5-8%", or something along these lines.
Revised this sentence.

3. Lines 72-73, I would omit the qualifier "broadly consistent", since the range (-43% to +133%) is rather large. I would suggest just stating the result of Turnbull et al (2011) as "were found to be within +X/-Y% of Vulcan".
Revised this sentence.

4. The β of equation (1) is not discussed in the text other than to say that it includes the influence of other terms like the biospheric disequilibrium flux. Is it assumed that β is perfectly

known? If so, that is fine, but that should be explicitly stated. Or, one could also given an estimate of β and say why it is unlikely to be a big factor for the estimates derived in the paper. Revised later sentence to state that uncertainty in β is included in total uncertainty.

5. Likewise, Δ_{bg} is not discussed after equation (1). I notice that the authors estimate a total emission outside of California in their inversion. Is this equivalent to estimating Δ_{bg}? Clarified that terms in Eq 1 are not calculated explicitly and Revised later sentence to state that background uncertainty is included in total uncertainty.

6. Lines 143-144, the four inventories mentioned cover different time periods. Are they normalized to the same California total before calculating the spread? If yes, then what determines the prior uncertainty of the California total ffCO2? If no, then isn't the spread artificially large because the inventories span different years? They are from different time periods but the differences between inventories are much larger than the differences between years (end of that paragraph). As there are differing trends in emissions between different emissions estimates for those available for individual years, it is not clear that normalizing to one year would reduce the spread.

7. Lines 146-147, was the standard deviation calculated across the four inventories, or was the spread (max to min) across four inventories assumed to be the 1-σuncertainty? The standard deviation was calculated across the four inventories, see updated lines 161-162 and Fischer et al., 2017.

8. Lines 184 and 205, the footprints are aggregated over six days, beyond day 1. Does this mean that the flux adjustments, beyond the first 24 hours, are all coherent across six days? Is that realistic? I'm curious why this was done, since I would assume the transport model would be able to distinguish between signals coming from flux 2 days ago vs 6 days ago (say). We simply mean that the impact of including diurnal emissions more than 24 hours before observation on the resulting concentration is not significant compared to invariant fluxes - these are findings by co-author Emily White from the University of Bristol. Added a statement to clarify "This approach captures the influence of temporally varying emissions that can be significant in the first 24 hours but we assume to be negligible for the period longer than 24 hours back in time."

9. Lines 272-273, saying "where there are regional errors in the magnitude of prior emissions" is not quite exact, I think. I suggest rephrasing this as "... in the simplest case where the only errors in prior regional flux estimates are biases in their magnitudes". Revised.

10. Lines 287-288, it's common practice in OSSE studies to use the more realistic scenario as the truth (nature run) and the simpler scenario as the prior. However, here the authors use

annually averaged Vulcan (less realistic) as the truth and temporally varying Vulcan (more realistic) as the prior. Why?

Revised to clarify in lines 318-323:

"It may seem counter intuitive to choose the simpler scenario (i.e. time invariant) as true emissions, however this was unfortunately due to the simulations available; we did not have simulated $ffCO_2$ concentrations from each air basin for temporally invariant emissions coupled with W-S-LBL footprints, only the total $ffCO_2$ concentrations. We do not expect that switching the prior and true emissions would significantly affect our conclusions."

11. Lines 289-290, I would have thought that annually averaged Vulcan would have the same total as temporally varying Vulcan, since averaging conserves the total. So why was scaling necessary? Was it because the inversions only covered a few months and not an entire year?

The inversion only covers periods in May 2014, October-November 2014 and January-February 2015 (see lines 122-124). May and October-November are troughs in the annual emission cycle, whilst January-February are peaks (see figure 1b).

12. Lines 298-299, similar question as before. The authors used annually averaged Vulcan (simpler scenario) as truth and prior instead of the more realistic temporally varying Vulcan. Why?

Please refer to the answer for point 10.

13. Lines 458-459. While I certainly understand the value of more observations, and am all for increasing the observations coverage of the $\Delta_{14}CO_2$ network, I do not think that having more observations will necessarily reduce the impact of transport model uncertainty. As the authors have themselves noted, the impact of transport model uncertainty is higher for smaller regions, while for larger regions (entire California) there is some cancellation. This is because the difference between transport models is typically more prominent at smaller scales (e.g., in the CO_2 inversion world, the global total flux is the easiest thing to estimate). So having more observations from a denser network could also sample these model differences even more and increase the impact of transport uncertainty on posterior flux estimates.

We think the word dense caused some confusion here and so we have revised this sentence to delete "dense" and to add the phrase "with multiple sites in a variety of settings" for clarification.

14. Lines around 490, and figures 3-6. The posterior bias is typically lower than the posterior uncertainty, barring a few exceptions. This could either be because the posterior biases are low (good outcome), or because the posterior uncertainties are large (less desirable outcome). Let's say that in an ideal world, we commit to making more $\Delta_{14}CO_2$ measurements of higher precision, which will reduce posterior uncertainty. Will that also decrease the biases in figures 3-6? Or will it increase some of the biases (see earlier point about transport uncertainty), and may

decrease others? Basically, what I'm trying to get at here is whether the good outcome for most of the flux estimates (bias < 2σ) is a happy accident of the specific 2018 network and measurement precision, or whether there is a more fundamental reason we can expect biases to be lower than posterior uncertainties under different (possibly increased) coverage scenarios. If you reduce uncertainty in observations, this will allow the inversion to scale emissions more (presuming prior uncertainty remains constant), hence transport bias could increase whilst posterior uncertainty decreases. Hypothetically this could drive the inversion to have a posterior bias that is larger than the associated 2σ posterior uncertainty. However we did conduct experiments (see response to Anonymous Referee) whereby we varied the transport error uncertainty parameter to 0.3 and 0.8 (hence changing the balance of prior - observation/transport uncertainty), with no significant differences in the result and all results were within 2σ of the truth.

15. Line 515, the Bagley et al (2017) reference is missing from the bibliography.
Fixed.

16. Line 534, suggest changing "much larger uncertainties" to "much larger percent- age uncertainties".
Corrected.

1 **Characterizing Uncertainties in Atmospheric Inversions of Fossil Fuel CO₂ Emissions in**

2 **California**

3 Kieran Brophy[1], Heather Graven[1], Alistair J. Manning[2], Emily White[3], Tim Arnold[4,5], Marc L
4 Fischer[6], Seongeun Jeong[6], Xinguang Cui[6], Matthew Rigby[3]
5
6 1. Department of Physics, Imperial College London, London, UK
7 2. Hadley Centre, Met Office, Exeter, UK
8 3. School of Chemistry, University of Bristol, Bristol, UK
9 4. National Physical Laboratory, London, UK
10 5. University of Edinburgh, Edinburgh, UK
11 6. Lawrence Berkeley National Laboratory, Berkeley, CA
12
13 Corresponding author:
14 Kieran Brophy
15 Email: kb613@ic.ac.uk

16 **Key Words:** Fossil Fuel, Carbon Dioxide, Simulation Experiment, Inversion, Transport Error

17 **Abstract**

18 Atmospheric inverse modelling has become an increasingly useful tool for evaluating emissions

19 of greenhouse gases including methane, nitrous oxide and synthetic gases such as

20 hydrofluorocarbons (HFCs). Atmospheric inversions for emissions of CO_2 from fossil fuel

21 combustion (ffCO₂) are currently being developed. The aim of this paper is to investigate

22 potential errors and uncertainties related to the spatial and temporal prior representation of

23 emissions and modelled atmospheric transport for the inversion of ffCO₂ emissions in the U.S.

24 state of California. We perform simulation experiments based on a network of ground-based

25 observations of CO_2 concentration and radiocarbon in CO_2 (a tracer of ffCO₂), combining prior

26 (bottom-up) emission models and transport models currently used in many atmospheric studies.

27 The potential effect of errors in the spatial and temporal distribution of prior emission estimates

28 is investigated in experiments by using perturbed versions of the emissions estimates used to

29 create the pseudo data. The potential effect of transport error was investigated by using three

30 different atmospheric transport models for the prior and pseudo data simulations. We find that

31 the magnitude of biases in posterior state-total emissions arising from errors in the spatial and

32 temporal distribution in prior emissions in these experiments are 1-15% of posterior state-total

33 emissions, and generally smaller than the 2-σ uncertainty in posterior emissions. Transport error

34 in these experiments introduces biases of -10% to +6% in posterior state-total emissions. Our

35 results indicate that uncertainties in posterior state-total $ffCO_2$ estimates arising from the choice

36 of prior emissions or atmospheric transport model are on the order of 15% or less for the ground-

37 based network in California we consider. We highlight the need for temporal variations to be

38 included in prior emissions, and for continuing efforts to evaluate and improve the representation

39 of atmospheric transport for regional $ffCO_2$ inversions.

40 **1. Introduction**

41 The U.S. state of California currently emits roughly 100 Tg C of fossil fuel CO_2 ($ffCO_2$) each

42 year (CARB, 2017), or approximately 1% of global emissions (Boden et al., 2016). The passing

43 of California's "Global Warming Solutions Act" (AB-32) in 2006 requires that overall

44 greenhouse gas emissions in California be reduced to their 1990 levels by 2020 (a 15% reduction

45 compared to business as usual emissions) with further reductions of 40% below 1990 levels

46 planned for 2030, and 80% below by 2050. The California Air Resources Board (CARB) is

47 responsible for developing and maintaining a "bottom-up" inventory of greenhouse gas

48 emissions to verify these reduction targets. However, previous studies have shown such

49 inventories may have errors or incomplete knowledge of sources (e.g. Marland et al, 1999;

50 Andres et al., 2012). Uncertainties in inventories of annual $ffCO_2$ emissions from most

51 developed countries (i.e. UNFCCC Annex I and Annex II) have been estimated to be between 5-

2

52 10% (Andres et al., 2012), and uncertainties can become much larger at subnational levels

53 (Hogue et al., 2016). In a recent study Fischer et al., (2017) found discrepancies between bottom-

54 up gridded inventories of $ffCO_2$ emissions were 11% of California's state total emissions.

55 Previous research has shown that inferring $ffCO_2$ emissions from atmospheric measurements,

56 including measurements of $ffCO_2$ tracers, could provide independent emissions estimates on

57 urban to continental scales (e.g. Basu et al., 2016; Lauvaux et al., 2016; Fischer et al., 2017;

58 Graven et al. 2018). Such estimates are derived from observations through the use of an

59 atmospheric chemical transport model and a suitable inverse method in a process often referred

60 to as "inverse modelling" or an "inversion". Distinguishing enhancements of CO_2 due to

61 anthropogenic or biogenic sources can be done using measurements of radiocarbon in CO_2

62 ($\Delta^{14}CO_2$), since CO_2 emitted from fossil fuel combustion is devoid of $^{14}CO_2$ due to radioactive

63 decay (Levin et al., 2003).

64 Recent studies with both real atmospheric measurements of $\Delta^{14}CO_2$ and with observing system

65 simulation experiments (OSSEs) at a network of sites have shown that atmospheric $\Delta^{14}CO_2$ can

66 be used to estimate monthly mean Californian $ffCO_2$ emissions with posterior uncertainties of 5-

67 8%, levels that are useful for the evaluation of bottom-up $ffCO_2$ emissions estimates.

68 Furthermore, Graven et al., 2018 found their posterior emissions estimates were not significantly

69 different from the California Air Resources Board's reported $ffCO_2$ emissions, providing

70 tentative validation of California's reported $ffCO_2$ emissions in 2014-15. In another study using

71 aircraft-based $\Delta^{14}CO_2$ measurements, Turnbull et al. (2011) found $ffCO_2$ emissions from

72 Sacramento County in February 2009 had a mean difference of -17%, range: -43% to +133%

73 with the Vulcan emissions estimate (Gurney et al., 2009).

3

76   Although atmospheric inversions may provide a method for estimating emissions that is useful

77   for evaluating emissions reduction policies, such as AB-32, systematic errors can arise from the

78   atmospheric transport and prior emission models (e.g., Nassar et al., 2014; Liu et al., 2014;

79   Hungershoefer et al., 2010; Chevallier et al., 2009, Gerbig et al., 2003). Comparisons of CO_2

80   simulated by different transport models have been conducted globally (e.g. Gurney et al., 2003,

81   Peylin et al. 2013), and on the European continental scale (Peylin et al., 2011). The latter found

82   that transport model error resulted in differences in modelled $ffCO_2$ concentrations that were 2-3

83   times larger than using the same transport model but different prior emissions, depending on the

84   location and time of year. However, comparisons of $ffCO_2$ simulated by different high resolution

85   models (25 km or less) at regional scales are still lacking.

86   The objective of this paper is to examine the sensitivity of a regional inversion for Californian

87    $ffCO_2$ emissions to errors in the prior emissions estimate and transport model. We build on

88   previous work by Fischer et al. (2017) that developed an Observation System Simulation

89   Experiment to estimate the uncertainties in both California statewide $ffCO_2$ emissions and

90   biospheric fluxes that might be obtained using an atmospheric inversion. Their inversion was

91   driven by a combination of in situ tower measurements, satellite column measurements from

92   OCO-2, prior flux estimates, a regional atmospheric transport modelling system, and estimated

93   uncertainties in prior CO_2 flux models, $ffCO_2$ measurements using radiocarbon, OCO-2

94   measurements, and in atmospheric transport. In contrast to Fischer et al., 2017 we focus only on

95    $ffCO_2$ emissions and use a network of flask samples without incorporating satellite

96   measurements.

97   Our approach is to use simulation experiments to quantify representation and transport error

98   using the inversion setup and the observation network from Graven et al. (2018) as a test case.

4

99   Specifically we test whether the inversion can estimate the "true" emissions that were used to

100 produce the pseudo data, within the uncertainties, when the prior emissions estimate includes

101 spatial and temporal representation errors within the scope of current emissions estimates

102 (Vulcan v2.2 and EDGAR v4.2 FT2010). We further test whether the inversion can estimate

103 "true" emissions, within the uncertainties, when the transport model used for the prior simulation

104 is different from the transport model used to produce the pseudo data, emulating transport error.

2. Data and Methods

106 The analysis approach applies a Bayesian inversion developed from previous work that combines

107 atmospheric observations, atmospheric transport modelling, prior flux models, and an

108 uncertainty specification (Jeong et al., 2013; Fischer et al., 2017). Here, the inversion scales prior

109 emission estimates in 15 regions (Figure 1a, Table 1) termed "air basins", classified by the

110 California Air Resources Board for air quality control

111 (https://www.arb.ca.gov/desig/adm/basincnty.htm).

2.1 Observation Network

113 As a test case to explore uncertainties in ffCO$_2$ inversions, we use the observation network of 9

114 tower sites in California that was used to collect flask samples for measurements of CO$_2$ and

115 radiocarbon in CO$_2$ in 2014-15 and simulate the same campaign periods (Figure 1a) (Graven et

116 al. 2018). Three, month-long, campaigns were conducted: 1st – 29th May 2014; 15th October –

117 14th November 2014; and 26th January – 15th February 2015, with flasks sampled approximately

118 every 2-3 days at 22:30 GMT (14:30 local standard time). We replicate the sample availability in

119 Graven et al. (2018), including the reduction in observation sites used in Jan-Feb 2015. The time

120 of observation was chosen as the planetary boundary layer is usually deepest in the afternoon so

5

121   that errors in the modelled boundary layer concentration are considered smaller (Jeong et al.,

122   2013), and afternoon concentrations are more representative of large regions.

123   The observed ffCO$_2$ concentration at a given site can be calculated by (Levin et al., 2003;

124   Turnbull et al. 2009):

125
$$ffCO_2 = C_{obs}\left(\frac{\Delta_{bg}-\Delta_{obs}}{\Delta_{bg}-\Delta_{ff}}\right) + \beta \qquad (1)$$

126   Where C_{obs} is the total observed CO_2 concentration at a given site. Δ refers to $\Delta^{14}C$, the ratio of

127    $^{14}C/C$ reported in part per thousand deviation from a standard ratio, including corrections for

128   mass-dependent isotopic fractionation and sample age (Stuiver and Polach, 1977). Δ_{bg}, Δ_{obs} and

129   Δ_{ff} are the $\Delta^{14}CO_2$ of background, observed and fossil fuel CO_2, respectively, where Δ_{ff} is -

130   1000‰ since ffCO$_2$ is devoid of $^{14}CO_2$. The term β is a correction for the effect of other

131   influences on $\Delta^{14}CO_2$, principally heterotrophic respiration (Turnbull et al. 2009). In the

132   experiments we present here, we do not explicitly calculate $\Delta^{14}CO_2$ or the other terms in

133   Equation 1, rather we simulate ffCO$_2$ and specify its uncertainty to be the same as the uncertainty

134   in radiocarbon-based estimates of ffCO$_2$. Following Fischer et al. (2017), total observational

135   uncertainty for ffCO$_2$ was assumed to be 1.5 ppm (1-σ), encapsulating measurement uncertainty,

136   background uncertainty and uncertainty in β. This is consistent with Graven et al. (2018), who

137   estimated total uncertainty in ffCO$_2$ for individual samples of 1.0 to 1.9 ppm.

138   **2.2 Prior Emissions Estimates and Prior Uncertainty**

139   The two prior emissions estimates used here are gridded products produced by EDGAR (version

140   FT2010) (EDGAR, 2011) for the year 2008 and Vulcan (version 2.2) for 2002 (Gurney et al.,

141   2009). EDGAR is produced at an annual resolution whilst Vulcan has an hourly resolution. The

6

142   two models use different emissions data and different methods to spatially allocate emissions

143   with annually averaged statewide emissions differing by 17.8 TgC (~19% of mean emissions),

144   and up to 11.6 TgC for individual air basins of California (Table 1). Although our campaigns

145   took place in 2014-2015, we use emissions estimates from Vulcan for the year 2002 and EDGAR

146   for 2008 as emissions estimates are not available from Vulcan and EDGAR for 2014-15. The

147   difference in state total emissions between 2002, 2008 and 2014-15 is 3-6 TgC (CARB, 2017),

148   much less than the EDGAR-Vulcan difference of 17.8 TgC.

149   We estimate prior uncertainty in the same way as in Fischer et al. (2017), using a comparison of

150   four gridded emissions estimates in California (Vulcan v2.2, EDGAR FT2010, ODIAC v2013

151   and FFDAS v2) as well as a comparison across an ensemble of emissions estimates for one

152   model (FFDAS v2, Asefi-Najafabady et al.,2014). Prior uncertainty is specified for the whole air

153   basin. The relative 1-σ standard deviation across the four inventories is between 8% and 100%

154   for individual air basins (Table 1), and this is what we use to specify the 1-σ uncertainty in the

155   prior emissions from each air basin. This estimate of prior uncertainty is referred to as "SD prior

156   uncertainty". We also conduct tests with an alternative prior uncertainty of 70% for each air

157   basin (referred to as "70% prior uncertainty"). This was done to test the sensitivity of our results

158   to the choice of prior uncertainty. Emissions occurring outside California were assumed to have

159   an uncertainty of 100% for both cases.

160   **2.3 Atmospheric Transport Models**

161   We simulate ffCO$_2$ using three different atmospheric transport models outlined in Table 2. These

162   models are commonly used in regional atmospheric transport modelling and greenhouse gas

163   inversion studies but to date have not been compared in California. Two of the transport models

7

164 use different versions and parameterizations of the Weather Research and Forecast (WRF) model

165 combined with the Stochastic Time-Inverted Lagrangian Transport (STILT) model. The third

166 transport model uses meteorology from the UK Met Office's Unified Model (UM) combined

167 with the Numerical Atmospheric dispersion Modelling Environment (NAME).

168 The first WRF-STILT model is run at Lawrence Berkeley National Laboratory (WS-LBL,

169 Fischer et al. 2017; Jeong et al. 2016; Bagley et al. 2017) and uses WRF version 3.5.1 (Lin,

170 2003; Nehrkorn et al., 2010). Estimates for Planetary Boundary Layer Height (PBLH) are based

171 on the Mellor–Yamada–Nakanishi–Niino version 2 (MYNN2) parameterization (Nakanishi and

172 Niino 2004, 2006). As in Jeong et al. (2016), Fischer et al. (2017) and Bagley et al. (2017), two

173 land surface models (LSMs) are used depending on the location of the observation site. A 5-layer

174 thermal diffusion land surface model is used in the Central Valley for the May campaign whilst

175 the Noah LSM (Chen et al., 2001) is used in the remaining campaigns and regions of California.

176 We implement multiple nested domains, with the outermost domain spanning 16-59°N and 154-

177 137°W with a 36km resolution, a second domain of 12km resolution over western North

178 America, and a third domain of 4km resolution over California. Two urban domains of 1.3 km

179 resolution are used in the San Francisco Bay area and the metropolitan area of Los Angeles.

180 Footprints describing the sensitivity of an observation to surface emissions are calculated by

181 simulating 500 model particles and tracking them backward for 7 days. The footprint of a given

182 site and observation time is produced hourly for particles below 0.5 times the PBLH.

183 The second WRF-STILT model is from CarbonTracker-Lagrange (WS-CTL), an effort led at

184 NOAA to produce standard footprints for greenhouse gas observation sites in North America

185 (https://www.esrl.noaa.gov/gmd/ccgg/carbontracker-lagrange). WS-CTL uses WRF version

186 2.1.2 and the Yonsei University (YSU) (Hong et al., 2006) PBLH scheme coupled with the Noah

8

187 land surface model and the MM5 (fifth generation Pennsylvania State University-National

188 Center for Atmospheric Research Mesoscale Model, Grell et al., 1994) similarity theory-based

189 surface layer scheme. As with WS-LBL, simulations are run for 7 days and particles below 0.5

190 times the PBLH are used in the calculation of the footprint. Footprints have a spatial resolution

191 of 0.1° for the first 24 hours and 1° for the remaining 6 days. Footprints are hourly dis-

192 aggregated for the first 24 hours and then aggregated for the remaining 6 days. This approach

193 captures the influence of temporally varying emissions that can be significant in the first 24

194 hours but we assume to be negligible for the period longer than 24 hours back in time. The 0.1°

195 spatial resolution domain is 31° longitude by 21° latitude with the domain centered on the release

196 location. The 1° resolution has a domain of 170°E to 50°E longitude and 10° N to 80° N latitude.

197 The WRF domain covers most of continental North America (Fig. 1 in Nehrkorn et al., 2010)

198 with 30 km resolution and has a finer nest with 10 km spatial resolution over the continental

199 United States. WS-CTL simulates footprints for 500 particles released over a 2-hour period

200 between 21:00 and 23:00 GMT (13:00 and 15:00 PST). An exception is Sutro Tower (STR),

201 where footprints are only available for an instantaneous release of 500 particles at 22:10 GMT.

202 Walnut Grove (WGC) footprints are available only for a release height of 30m a.g.l, which is

203 lower than the sampling height of 91m a.g.l. used in the observation campaign (Graven et al.

204 2018) and used in the other two transport models. Footprints were available for 2014 but not for

205 2015, so the WS-CTL model is used for simulations of the May and Oct-Nov 2014 campaigns

206 but not for the Jan-Feb 2015 campaign.

207 The third model, UM-NAME, is the UK Met Office's NAME model, Version 3.6.5 (Jones et al.,

208 2007), driven by meteorology from the Met Office's global numerical weather prediction model,

209 the Unified Model (UM) (Cullen et al., 1993). The UM model has a horizontal resolution of ~25

9

210   km up to July 2014, covering the period of the May 2014 campaign. The horizontal resolution

211   was then increased to ~17 km covering both the October-November 2014 and January-February

212   2015 campaigns. The temporal resolution of the UM meteorology is every 3 hours for all

213   campaigns. Following a similar approach as for the WRF-STILT models, 500 particles were

214   released instantaneously at 22:30 GMT and simulated for hourly dis-aggregated footprints for the

215   first 24 hours and aggregated for the remaining 6 days. The footprints are calculated for the same

216   horizontal resolution as the UM meteorology (25 or 17km), where the particles present in the

217   layer between 0 and 100 m above ground level are used to calculate the footprint. The

218   computational domain covers 175.0°W to 75°W longitude and 6.0°N to 74°N latitude.

219   Simulated ffCO$_2$ signals (the enhancement of CO$_2$ concentration due to ffCO$_2$ emissions within

220   the model domain) are calculated by taking the product of the footprint and an emissions

221   estimate, both with the spatial resolution of the footprint at the native footprint resolution.

222   The resulting concentration is summed for individual air basins. Following previous work,

223   we assume a transport model uncertainty of 0.5 times the mean monthly signal in the pseudo-

224   observations at each site (referred to as the 'uncertainty parameter') (Jeong et al., 2013; Fischer

225   et al 2017). We also test the effect of changing the uncertainty parameter to 0.3 and 0.8.

226   Ten ensembles were run for UM-NAME to test the effect of random errors on the calculation of

227   the footprint. The RMSE was within 10% of the mean monthly signal for most observation sites.

228   This is similar to the findings of Jeong et al. (2012), which the transport model uncertainty is

229   based on. Two observation sites (THD and VTR) had slightly higher RMSE, but both were

230   within 20% of the mean monthly signal.

231   **2.4 Inversion Method**

232 Our inversion method is a Bayesian synthesis inversion to scale emissions in separate regions of

233 California. We follow the same approach as Fischer et al. (2017) to solve for a vector of scaling

234 factors, λ, for 15 air basins and a 16th region representing the area outside of California. Unlike

235 Fischer et al. (2017), we do not split the San Joaquin Valley into two regions. The inversion uses

236 the set of observations, c, and the matrix of predicted ffCO$_2$ signals from each air basin, K, to

237 optimize the cost function J:

238 $$J_\lambda = (c - K\lambda)^T R^{-1}(c - K\lambda) + (\lambda - \lambda_{prior})^T Q_\lambda^{-1}(\lambda - \lambda_{prior}) \quad (2)$$

239 λ$_{prior}$ is the prior estimate of the scaling factors (a vector of ones with length equal to the number

240 of regions) and R and Q$_\lambda$ are the error covariance matrices relating to observational and model

241 transport errors, and prior emissions estimate errors respectively. The non-diagonal elements of

242 R and Q$_\lambda$ are zero, assuming uncorrelated errors in the prior emissions in each air basin and in

243 the model and observations. This assumption for R is robust as we only generate one pseudo

244 observation every 2-3 days. Included in R are observational errors and transport model errors,

245 added in quadrature. Therefore if the average signal at an observation site is very small, then

246 observational uncertainty (1.5 ppm) will dominate R. Minimizing J using the standard least

247 squares formulation under the assumption of Gaussian distributed uncertainties gives the

248 posterior estimate for λ following:

249 $$\boldsymbol{\lambda}_{post} = \left(\mathbf{K}^T\mathbf{R}^{-1}\mathbf{K} + \mathbf{Q}_\lambda^{-1}\right)^{-1}\left(\mathbf{K}^T\mathbf{R}^{-1}\mathbf{c} + \mathbf{Q}_\lambda^{-1}\boldsymbol{\lambda}_{prior}\right) \quad (3)$$

250 With the posterior error covariance given as:

251 $$V_{post} = \left(\mathbf{K}^T\mathbf{R}^{-1}\mathbf{K} + \mathbf{Q}_\lambda^{-1}\right)^{-1} \quad (4)$$

252 λ$_{post}$ and V$_{post}$ are computed separately for each of the three campaigns outlined in section 2.1.

253 Posterior emissions estimates are the product of λ$_{post}$ and the prior emissions estimate from each

254 air basin. State total emissions are then calculated by summing the emissions in each air basin.

11

255 Uncertainty in the state-wide Californian posterior flux, including error correlations, is calculated

256 as:

257 $$\sigma_E{}^2 = E_{prior} V_{post} E_{prior}^T \qquad (5)$$

258 Where E_{prior} is a vector of $ffCO_2$ emissions from each air basin.

259 **2.5 Simulation Experiments**

260 We conduct a series of experiments to test the performance of the inversion in estimating the true

261 emissions when the emissions estimates or transport models used to produce pseudo-

262 observations are different to those used to produce the prior simulations. The tests explore the

263 effect differences in the magnitude, spatial distribution, and temporal variation of prior emissions

264 have on posterior emissions. We also examine the effect of using different transport models to

265 simulate pseudo observations and to simulate prior concentrations.

266 As part of these experiments, we evaluate the impact of outlier removal on the simulation

267 experiments. Outlier removal is generally used in atmospheric inversions when there is an issue

268 with the ability of the model to simulate a particular observation. We use the outlier removal

269 method outlined in Graven et al. (2018) and compare with inversion results where no outliers are

270 removed. In this outlier removal method, an observation (here, a pseudo-observation) is

271 designated as an outlier if (1) the absolute difference between the $ffCO_2$ signals in the

272 observation and the prior simulation is greater than the average of the observed and simulated

273 $ffCO_2$, and (2) either the observed or simulated $ffCO_2$ is greater than 5 ppm.

274 **2.5.1 Difference in magnitude of emissions**

275 First we test how well the inversion estimates the true emissions if the prior emissions have a

276 systematic error in magnitude, but no error in the spatial or temporal distribution of emissions

12

277　and no error in atmospheric transport. In this experiment, the prior emissions estimate is given by

278　EDGAR and the true ffCO$_2$ signals were generated by scaling the EDGAR emissions in each air

279　basin to match the annually averaged Vulcan emissions in that air basin. These differences range

280　from 0.1 TgC in San Diego to 11.6 TgC in the San Joaquin Valley (Table 1). The EDGAR state

281　total emissions are 12% higher than Vulcan, so the bias in the prior estimate in the state total

282　ffCO$_2$ emissions is +12%. The experiment is run for all the transport models with no temporal

283　variation in emissions. This experiment assesses the performance of the inversion and the

284　strength of the data constraint provided by the observation network in the simplest case where

285　the only errors in prior regional flux estimates are biases in their magnitudes. Prior uncertainty

286　is fixed per air basin for all experiments.

287　**2.5.2 Difference in spatial distribution of emissions**

288　To investigate the bias in the posterior emissions estimate that could result from errors in the

289　spatial distribution of prior emissions within each air basin, we now use annually averaged

290　Vulcan emissions as the true emissions and EDGAR emissions scaled in each air basin to match

291　the annually averaged Vulcan emissions in that region as the prior estimate of emissions. In this

292　experiment, the prior estimate of the total emissions in each air basin is unbiased, and we assess

293　how differences in the spatial distribution of emissions between Vulcan and EDGAR in each air

294　basin may lead to a bias in the posterior emissions estimate. As shown in Figure 1c, the most

295　significant discrepancies in spatial distribution are in the major urban areas of Los Angeles and

296　the San Francisco Bay. This experiment is also run for all the transport models using the same

297　transport model for both the true and prior simulation and including no temporal variation in

298　emissions.

13

299 **2.5.3 Difference in temporal variation of emissions**

300 To assess the impact of temporally-varying emissions on the inversion result, we generated true

301 ffCO$_2$ signals with temporally-invariant annually-averaged Vulcan emissions and prior ffCO$_2$

302 signals with temporally-varying Vulcan emissions. It may seem counter intuitive to choose the

303 simpler scenario (i.e. time invariant) as true emissions, however this was dictated by the

304 simulations available; we did not have simulated ffCO$_2$ concentrations from each air basin for

305 temporally invariant emissions coupled with W-S-LBL footprints, only the total ffCO$_2$

306 concentrations. We do not expect that switching the prior and true emissions would significantly

307 affect our conclusions. We scaled the temporally-varying Vulcan emissions in each air basin so

308 that the total ffCO$_2$ emissions were the same magnitude as the total ffCO$_2$ emissions in the

309 annually averaged Vulcan emissions for each field campaign. As shown in Figure 1d, scaling

310 was less than 10% of annual mean emissions with campaigns occurring during maxima and

311 minima of the annual emissions cycle. Here the prior estimate is again unbiased, and we assess

312 how differences in the diurnal variation of emissions (see Fig 1b) may lead to a bias in the

313 posterior emissions estimate. This experiment is also run for all the transport models using the

314 same transport model for both the true and prior simulation. Prior uncertainty is specified relative

315 to prior emissions, hence it differs in absolute magnitude for monthly differences in emissions.

316 Over the state this variation is ~15% when comparing May/Oct-Nov to Jan-Feb (see Fig. 1d).

317 **2.5.4 Difference in Atmospheric Transport**

318 To test the effect of differences in the simulated atmospheric transport of emissions, the same

319 emissions estimate (annually-averaged Vulcan) is coupled with two different transport models to

320 generate prior and true ffCO$_2$ signals. This experiment investigates potential effects of transport

14

321 errors, within the variations in transport across the three models we use. WS-LBL is considered

322 the "true" atmospheric transport while UM-NAME and WS-CTL are used for the prior

323 simulation in individual experiments. Here the prior estimate is again unbiased, and we assess

324 how differences in the modeled atmospheric transport may lead to a bias in the posterior

325 emissions estimate.

326 **3 Results**

327 **3.1 Simulated ffCO$_2$ Observations**

328 Before presenting the results of the inversion experiments, we first examine simulated ffCO$_2$

329 contributions different regions at each of the 9 observation sites. This allows us to quantify

330 which air basins have the largest influence on simulated concentrations at observation sites and

331 better interpret the results of the experiments. Figure 2 shows simulated concentrations at

332 observation sites resulting from emissions in the 6 highest-emitting air basins in California, and

333 from outside California. The highest signals (> 10 ppm) are simulated at urban sites (e.g. CIT

334 and SBC) for emissions from urban air basins (e.g., South Coast, 14.SC). The 9 air basins not

335 shown in Fig. 2 contributed signals below 0.1 ppm due to the small size or low emissions of the

336 air basin (e.g. Lake County and Lake Tahoe), or distance from the observation network (e.g.

337 Northeast Plateau, Great Basin Valleys and Salton Sea). In general, the northern sites (THD to

338 SLT in Fig 2) are sensitive to northern air basins (Sacramento and San Joaquin Valleys and SF

339 Bay), and the southern sites (VTR to SIO) are sensitive to emissions from southern air basins

340 (Mojave Desert, South Coast and San Diego). All transport models show the observation sites

341 are sensitive to more air basins in the Oct-Nov and Jan-Feb campaigns, compared to the May

15

342   campaign (Fig. 2). Signals simulated by WS-CTL come from fewer air basins than UM-NAME

343   or WS-LBL, particularly in May.

344   In our simulation experiments, signals from outside California are generally small compared to

345   the total signal for most sites (<10% on average), although they can average 20-50% for STB,

346   STR, SLT and SIO for individual campaigns. For THD, located near the northern border of the

347   state, a larger influence from outside California is found, 10-90%, due to a combination of

348   relatively low local emissions and northerly winds transporting emissions from the northwestern

349   United States and Canada.

350   **3.2.1 Difference in magnitude of emissions**

351   Figure 3 (a) shows the statewide inversion result for the experiment testing the effect of a bias in

352   magnitude in regional emissions in the prior simulation. In this figure, and similar figures that

353   follow for the other experiments, prior estimates are represented by black markers and posterior

354   estimates are represented by colored markers, with the 2-σ uncertainty on the x-axis and the bias

355   relative to the truth on the y-axis. The diagonal lines show 1:1 and 1:-1 lines, so that a marker

356   lying to the right of these lines indicate the posterior bias is smaller than the posterior

357   uncertainty, whereas a marker to the left of these lines indicate the posterior bias is larger than

358   the posterior uncertainty. Filled markers show results using SD prior uncertainty and empty

359   markers show results using 70% prior uncertainty. Prior and posterior uncertainties are expressed

360   as 2-σ.

361   For all transport models and campaigns, the inversion is able to reduce prior bias and scale

362   posterior emissions towards the truth. The +12% bias in the statewide emissions in the prior was

363   reduced to a posterior bias of between 0 and +9% (mean bias = +5%) for SD prior uncertainty.

16

364   Using 70% prior uncertainty reduced prior bias to between -3 and +6 (mean = +1%). Statewide

365   posterior uncertainty was 10-14% (mean 12%) and 14-32% (mean = 21%) for SD and 70% prior

366   uncertainty respectively, where uncertainty is expressed as 2-σ, lower than the statewide prior

367   uncertainties of 16% for SD and 69% for 70% prior uncertainty. There were no outliers

368   identified in this experiment.

369   To determine what is driving the statewide results, we examine the individual air basin inversion

370   results. Figure 3 (b) shows the inversion results for the six main emission regions of California,

371   with San Joaquin Valley (8.SJV) and South Coast (14.SC) having the largest prior biases. For the

372   San Joaquin Valley (8.SJV) and South Coast (14.SC) regions with the largest prior bias, the

373   biases are reduced in most cases, however, only the posterior estimates from the 70% prior

374   uncertainty experiment overlap the true emissions. The posterior estimates for SD prior

375   uncertainty do not overlap with the truth, indicating that the 2-σ prior uncertainty of 24% in

376   South Coast (14.SC), for example, restricts the inversion from eliminating biases of 30% in these

377   regions (Table 1), given the observations available. The 9 air basins omitted from Fig. 3(b) are

378   generally not being scaled by the inversion due to a lack of constraint from the observation

379   network, low emissions, or small prior uncertainty (Figure S1).

380   The bias in the posterior estimate of statewide emissions is larger in May than in Oct-Nov and

381   Jan-Feb (Fig 3a, triangles). This poorer performance of the inversion in May can be largely

382   attributed to the San Joaquin Valley (8.SJV), where the posterior emissions are largely

383   unchanged from the prior in May. There is no observation site in the San Joaquin Valley, and as

384   shown in Fig. 2, emissions in the San Joaquin Valley do not reach observation sites in

385   neighboring air basins in May, but they do reach these sites in Oct-Nov and Jan-Feb. In contrast,

386   the South Coast (14.SC) influences the two observation sites, CIT and SBC, located in the region

387   as well as several other sites (Fig. 2). Both CIT and SBC show prior signals are too high

388   compared to true signals for all campaigns and models (Fig. 3c), reflecting the positive bias in

389   prior emissions in the South Coast region, which is reduced in the posterior. Changing the

390   uncertainty parameter from 0.5 to 0.3 or 0.8 had the result of decreasing the ability of the

391   inversion to scale state-wide emissions towards true emissions by 1-4%, with an increase in

392   posterior uncertainty by a similar percentage.

393   **3.2.2 Difference in spatial distribution of emissions**

394   The statewide inversion results for the experiment including errors in the spatial distribution of

395   emissions are shown in Figure 4 (a). In this case the magnitude of prior emissions in each air

396   basin is equal to true emissions and we aim to quantify how errors in the spatial distribution of

397   emissions (EDGAR as prior and Vulcan as true distribution) lead to bias in posterior emissions

398   estimates. Posterior emissions are negatively biased, apart from WS-LBL in January-February.

399   Posterior bias was between -10% and +1% (mean -4%) for SD prior uncertainty and between -

400   10% and +4% (mean = -4%) for 70% prior uncertainty across transport models and campaigns.

401   As might be expected from the experimental setup with an unbiased prior, posterior emissions

402   estimates generated using SD prior uncertainty have a smaller mean bias and smaller range of

403   posterior estimates compared to those generated using 70% prior uncertainty. Statewide

404   uncertainty was reduced from 16% to 10-14% (mean = 12%) for SD prior uncertainty and from

405   58% to 14-21% (mean = 18%) for 70% air basin prior uncertainty. Biases induced are smaller

406   than the 2-σ posterior uncertainty across all transport models, campaigns and choice of prior

407   uncertainty.

408   Posterior emissions results in the two largest emitting air basins (the San Francisco Bay and

409   South Coast) are also negatively biased in most cases (Fig 4b). In several cases, posterior biases

410   are larger than the associated posterior uncertainties, for example in the South Coast for WS-

411   LBL in all cases. Considering Figure 4 (c), prior ffCO$_2$ signals are being overestimated more

412   often than underestimated, particularly for the relatively more urban sites CIT and SLT.

413   Since the prior emissions from EDGAR have been scaled to have the same total as Vulcan (the

414   true emissions) in each region, the pattern of more negative posterior emissions is only caused by

415   the sub-regional spatial distribution of emissions. Comparing Vulcan and EDGAR native grid

416   cell emissions in Figures 1c and S2, EDGAR tends to have greater emissions in high-emission

417   grid cells. In other words, the emissions are more concentrated in EDGAR and more dispersed in

418   Vulcan. This pattern explains the negative bias in posterior emissions for the urban South Coast

419   air basin. The opposite effect does not appear to hold for rural observation sites and regions,

420   perhaps because rural emissions are already rather dispersed and have less of an influence on the

421   observations.

422   In these experiments, 0-3% of observations were identified as outliers, but excluding outliers did

423   not change the statewide result significantly (<1% change in mean bias).

424   **3.2.3 Difference in temporal variation of emissions**

425   Figure 5 (a) shows the statewide inversion result for the experiment where the emissions are

426   Vulcan temporally-varying in the prior simulation (see Fig. 1b) but Vulcan temporally-invariant

427   in the true simulation. Posterior bias was between -13 and +5% (mean = -3%) for SD uncertainty

428   and between -15% and +6% (mean = -3%) for 70% prior uncertainty. Posterior uncertainty was

429   11-15% (mean = 12%) for SD prior uncertainty and 15-24% (mean = 18%) in posterior

Deleted: Sacramento Valley (3.SV) and the San Joaquin Valley (8.SJV) have higher posterior emissions in WS-LBL in most cases, possibly due to the inversion compensating for reduced posterior emissions in the San Francisco Bay (13.SFB) and South Coast (14.SC).

19

435   emissions for SD (70%) prior uncertainty. Outlier removal resulted in 0-1% (mean = 0%) of data

436   points being removed, which did not change the statewide results.

437   The posterior estimate for WS-LBL in May with SD prior uncertainty has a significant negative

438   bias of -13%, approximately the same magnitude as the associated 2-σ posterior uncertainty. As

439   can be seen by the air basin results of Figure 5 (b), the statewide bias for WS-LBL in May is

440   being driven by a large regional bias in the South Coast, but also in the San Francisco Bay and

441   San Diego air basins. These regional biases are larger than their associated posterior

442   uncertainties. Figure 5 (c) shows the prior ffCO$_2$ signals at CIT average ~7ppm too high in May

443   for WS-LBL. In contrast, prior ffCO$_2$ signals at CIT and SBC are too low in Oct-Nov for WS-

444   CTL, leading to a high bias in posterior emissions from the South Coast. San Diego also

445   exhibited both high and low biases in the posterior emissions. Overall, temporal variations in

446   emissions led to posterior biases generally within ±6%, but as large as 15%; however, a

447   consistent pattern in the posterior bias due to the temporal representation in emissions was not

448   found.

449   **3.2.4 Difference in Atmospheric Transport**

450   The statewide inversion results for the experiment where the atmospheric transport in the prior

451   simulation uses WS-CTL or UM-NAME but the atmospheric transport in the true simulation

452   uses WS-LBL are shown in Figure 6 (a). Outliers were identified in these experiments and we

453   present results for inversions including all data and for inversions where outliers were removed.

454   When all data are included, differences in atmospheric transport model introduces a bias in

455   statewide posterior emissions of between -42% and -3% (mean = -12%) for SD prior uncertainty

456   and between -32% and 0% (mean = -15%) for 70% prior uncertainty. For one case, using WS-

20

457   CTL to generate prior signals in October-November, the bias in the posterior emissions estimate

458   was larger than the 2-σ uncertainty for both SD and 70% prior uncertainty. Changing the

459   uncertainty parameter from 0.5 to 0.3 or 0.8 resulted in posterior emissions remaining closer

460   to true emissions by 0-4% and increasing the posterior uncertainty by 1-5%.

461   Removing outliers significantly improved the inversion results (Figure 6 b): the mean bias was

462   between -10% and 0% (mean = -3%) for SD prior uncertainty and between -9% and +6% (mean

463   = -5%) for 70% prior uncertainty when outliers were removed. Posterior uncertainty was 9-15%

464   (mean = 12%) and 15-24% (mean = 18%) for SD and 70% prior uncertainty respectively, with

465   all posterior estimates within 2-σ of the true statewide emissions. The reduction in posterior bias

466   when outliers are removed is mostly due to the removal of a few large positive outliers in prior

467   simulated signals by WS-CTL (Figure 7). Figure 7 illustrates the time series of simulated $ffCO_2$

468   in each model with outliers shown as an x. Outliers removed were between 6.9% and 20.6% of

469   all observations (mean = 10.5%). This is similar to the fraction of outliers identified in Graven et

470   al. 2018 using the same method with real data (~8%). It is also similar to that of Jeong et al.,

471   2012a and b (0-27%) for monthly inversions of CH_4 in California using a different method of

472   identifying outliers where model-data residuals are larger than 3-σ of model-data uncertainty.

473   This is an important result for the atmospheric inversion community working at such spatial

474   scales, as it highlights the benefits of removing outliers.

475   While the statewide posterior emissions estimate is significantly biased in only one case (WS-

476   CTL in Oct-Nov) when outliers are not removed, the posterior emissions estimates for the main

477   emissions regions are significantly biased in several cases (Fig 6c). The largest bias is in the

478   South Coast region where posterior estimates are biased by more than -75% (with 1% posterior

479   uncertainty) in Oct-Nov when using WS-CTL to generate prior signals. This large posterior

480 emissions bias in the South Coast and the statewide total can be attributed to overestimates in

481 prior ffCO$_2$ signal of more than 6ppm on average at CIT and SBC and more than 2ppm at WGC

482 and STR (Fig. 6e) due to some high outliers in the WS-CTL simulations (Fig. 7). Posterior

483 estimates for San Francisco Bay, South Coast and San Diego were also significantly biased in

484 some other cases, particularly for 70% prior uncertainty but also for SD prior uncertainty. This

485 indicates that regional biases caused by differences in atmospheric transport appear to

486 compensate over the statewide scale, and that results for individual regions are less robust than

487 aggregate results for the statewide network. It also suggests that an observation network with

488 multiple sites in a variety of settings is beneficial to reducing the impact of uncertainty in

489 atmospheric transport.

490 To investigate the differences in simulated ffCO$_2$ and assess whether these could be attributed to

491 specific aspects of modelled meteorology, we compared PBLH and wind speed in WS-LBL and

492 the UM for 5 of the 9 observation sites where PBLH output was available. PBLH was not

493 available for WS-CTL. Estimates for PBLH in WS-LBL are based on the Mellor–Yamada–

494 Nakanishi–Niino version 2 (MYNN2) parameterization scheme that estimates PBLH using

495 localized turbulence kinetic energy closure parameterization (Nakanishi and Niino 2004, 2006).

496 Estimates of PBLH are calculated internally within the UM. PBLH and wind speed were

497 averaged over 6 hours from 12 to 6pm Pacific Standard Time to compare the afternoon means

498 (Seibert et al., 2000). We found no consistent correlation between differences in PBLH or wind

499 speed and differences in simulated ffCO$_2$ between models across sites and campaigns (Figure

500 S3). Absolute values of wind direction and ffCO$_2$ did not show consistent correlations either. The

501 lack of correlation suggests we cannot attribute differences in simulated ffCO$_2$ to any single

502 meteorological variable estimated at any individual station in the transport models.

22

503 We also examined if differences in simulated $ffCO_2$ signals across transport models could be

504 explained by the differences in spatial resolution of the models. WS-CTL footprints were re-

505 gridded from a 0.1° native grid to the coarser UM-NAME grid of 17 or 25km and then used to

506 simulate $ffCO_2$. For this comparison, we simulated $ffCO_2$ every day over the campaign period.

507 We found no consistent reduction in mean $ffCO_2$ bias between sites over the 2 campaigns,

508 however there is a reduction in spread of bias at 4 sites for both campaigns (WGC, SLT, SBC

509 and SIO), suggesting model resolution could potentially have an impact for these sites. In general

510 however, we cannot say that transport model resolution error in atmospheric transport is a key

511 driver of $ffCO_2$ signal bias across observation sites (Figure S4).

512 **4 Discussion**

513 Our results show that atmospheric inversions can reduce a hypothetical bias in the magnitude of

514 prior $ffCO_2$ emissions estimates for the U.S. state of California using the ground-based

515 observation network from Graven et al. (2018), under the idealized assumptions of perfect

516 atmospheric transport and perfect spatio-temporal distribution of emissions in the prior estimate.

517 By exploring differences in model transport and spatio-temporal distribution of prior emissions,

518 we found that biases of magnitude 1-15% in monthly posterior estimates of statewide emissions

519 can result from differences in the temporal variation, spatial distribution and modelled transport

520 of the prior simulation. However, these biases were less than the 2-σ posterior uncertainty in

521 state-total emissions, when outliers were removed. In some cases, the biases in posterior

522 emissions for individual air basins were significant, compared to the posterior uncertainties,

523 suggesting that estimates for individual regions are less reliable than the aggregate estimates of

524 the state-wide total.

23

525 The largest bias in statewide posterior estimates was found to be caused by errors in the temporal

526 variation in emissions. This highlights the necessity for temporally-varying emissions to be

527 estimated and included in prior emissions estimates, particularly for urban regions. Similar

528 results have been found in other regions including Indianapolis (Turnbull et al. 2015) and Europe

529 (Peylin et al. 2011), and more generally, for high-emission regions around the globe (Zhang et al.

530 2016). Although the afternoon sampling is near to the diurnal maximum in emissions in

531 California (Fig. 1c, Gurney et al. 2009), which might be expected to lead to higher simulated

532 $ffCO_2$ in temporally-varying vs temporally-invariant emissions, we did not find consistently

533 positive biases in $ffCO_2$ but rather both positive and negative biases. This suggests the overall

534 impact of temporally-varying emissions depends on the model transport and the characteristics of

535 the observation site. Furthermore, uncertainties in the temporal distribution of emissions at an

536 hourly resolution have not yet been fully quantified (Nassar et al., 2013).

537 Errors in model transport, as represented in our experiments by using different transport models,

538 were shown to bias posterior $ffCO_2$ emissions by 10% or less, when outliers were removed.

539 These biases related to transport error are somewhat lower than estimated by similar simulation

540 experiments for $ffCO_2$ emissions estimates for the U.S. by Basu et al. (2016) using different

541 transport models (>10%), although their spatial scale was larger and the alternate model they

542 used was intentionally biased. In contrast, the three models we use are all actively applied in

543 regional greenhouse gas inversions. Our results are comparable to the estimate of ±15%

544 uncertainty in atmospheric transport in WS-LBL using comparisons with atmospheric

545 observations of CO in California (Bagley et al. 2017).

546 The fraction of pseudo-observations we identified as outliers in these transport error experiments

547 (10.5%, range 6.9-20.6%), was similar to Graven et al., 2018, where 8% of all observations were

24

548 removed as outliers using the same method. The outliers in our experiments were primarily high

549 ffCO$_2$ signals simulated by WS-CTL in Oct-Nov. When included in the inversion, these did lead

550 to significant biases in the posterior estimates for the experiment on model transport. This

551 highlights the need for careful examination of simulated ffCO$_2$ and consideration of outliers in

552 atmospheric ffCO$_2$ inversions.

553 Attributing differences in simulated ffCO$_2$ between the different transport models to specific

554 meteorological variables proved inconclusive, and model resolution error did not appear to

555 explain the differences in simulated signals, although we were not able to investigate aggregation

556 error in comparison to the high-resolution WS-LBL model. Wang et al. (2017) found

557 aggregation error to be only a minor contributor to errors in simulated ffCO$_2$ in Europe, while

558 Feng et al., (2016) found that high-resolution gridded emissions estimates could be more

559 important than high resolution transport models for simulations of greenhouse gases in Greater

560 Los Angeles. We found that differences in the spatial representation of prior emissions in

561 EDGAR compared to Vulcan led to consistently lower, although not significantly different,

562 posterior state-wide estimates due to the emissions in EDGAR being more concentrated in urban

563 regions. The spatial allocation of emissions between urban and rural regions in gridded emissions

564 estimates have much larger percentage uncertainties than national totals (Hogue et al. 2016),

565 suggesting that several different gridded emissions estimates should be used in regional ffCO$_2$

566 inversions to capture this source of uncertainty.

567 The results of these experiments suggest that the choice of prior emissions estimate and transport

568 model (among those considered here and currently used in the community) used in our ffCO$_2$

569 inversion would result in differences of 15% or less in posterior state-wide ffCO$_2$ emissions in

570 California, using the observation network from Graven et al. (2018). These differences are

571   generally not significant, compared to the posterior 2-σ uncertainties of 10 to 15%. In

572   comparison, Graven et al. (2018) found that posterior state-wide ffCO$_2$ emissions were not

573   statistically different when using temporally-varying emissions from Vulcan, as compared to

574   annual mean emissions from Vulcan or EDGAR, with posterior uncertainties of ±15 to ±17%.

575   Our results may be specific to the California region, observation network and inversion setup we

576   consider here, but we expect that similar differences of 1-15% are likely to be found elsewhere in

577   similar inversions at comparable regional scales. We note that while we have assessed individual

578   contributions to uncertainty in the experiments formulated here, these contributions can also

579   interact with each other. These interactions could act to increase the resulting biases, or partly

580   cancel them, depending on the combination used. The possibility for interacting effects implies

581   that multiple prior emissions estimates and transport models should be used in inversions of real

582   data.

583   In our results, emissions from many small or rural air basins did not have a significant

584   contribution to the local enhancement of ffCO$_2$ at the observation sites and were not adjusted by

585   the inversion in most cases (Figure 2, Figure S1). Combined with our experimental setup

586   specifying the magnitude of prior emissions to be equal to true emissions, it might be expected

587   that our results could underestimate the predicted biases in posterior emissions. However, these

588   experiments were designed specifically to quantify representation and transport error using the

589   inversion setup and the observation network from Graven et al. (2018) as a test case. Here, we

590   have assumed the model-measurement mismatch uncertainty matrix is diagonal, following

591   previous work (e.g. Gerbig et al. 2003; Fischer et al., 2017), however the consideration of

592   correlated errors in the uncertainty matrix has also been found to affect posterior emissions for

593   methane in California and reduce their uncertainty at the level of several percent (Jeong et al.

594   2016). Fischer et al. (2017), showed in individual simulation experiments that using either

595   EDGAR, or a spatially uniform flux of 1 μ mol m^{-2} s^{-1} as a biased prior, produced posterior

596   emissions that are substantially closer to true emissions, but only if the prior uncertainties are set

597   large enough to encompass biases in prior emissions. Therefore, further experiments using a

598   different experimental setup such as choice of mismatch error or specification of inversion

599   regions (e.g. to change the inversion region size based on proximity to the observation network,

600   Manning et al., 2011), would help to characterize uncertainties in regional ffCO$_2$ inversions and

601   the robustness of posterior estimates to the choices made in the inversion setup.

602   **Conclusion**

603   We have shown that atmospheric inversions for the U.S. state of California can reduce a

604   hypothetical bias in the magnitude of prior emissions estimates of ffCO$_2$ in California using the

605   ground-based observation network from Graven et al. (2018). Experiments to characterize the

606   effect of differences in the spatial and temporal distribution in prior emissions resulted in biases

607   in posterior state-total emissions with magnitudes of 1-15%, similar to monthly posterior

608   estimates of Basu et al., 2016 for the western United States. Our results highlight the need for (1)

609   temporal variation to be included in prior emissions, (2) different estimates of the spatial

610   distribution of emissions between urban and rural regions to be considered, and 3) representation

611   of atmospheric transport in regional ffCO$_2$ inversions to be further evaluated.

27

References

621

622 Andres, R. J., Boden, T. A., Bréon, F. M., Ciais, P., Davis, S., Erickson, D., ... & Oda, T. (2012).

623   A synthesis of carbon dioxide emissions from fossil-fuel combustion. Biogeosciences, 9(5),

624   1845-1871.

625

626 Asefi-Najafabady, S., Rayner, P. J., Gurney, K. R., McRobert, A., Song, Y., Coltin, K., ... &

627   Baugh, K. (2014). A multiyear, global gridded fossil fuel CO_2 emission data product:

628   Evaluation and analysis of results. Journal of Geophysical Research: Atmospheres, 119(17),

629   10-213.

630

631 Bagley, J. E., Jeong, S., Cui, X., Newman, S., Zhang, J., Priest, C., ... & Lareau, N. (2017).

632   Assessment of an atmospheric transport model for annual inverse estimates of California

633   greenhouse gas emissions. Journal of Geophysical Research: Atmospheres, 122(3), 1901-

634   1918.

635

636 Basu, S., Miller, J. B., & Lehman, S. (2016). Separation of biospheric and fossil fuel fluxes of

637   CO_2 by atmospheric inversion of CO_2 and $^{14}CO_2$ measurements: Observation System

638   Simulations. Atmospheric Chemistry and Physics, 16(9).

639

640   Boden, T., Andres, B., & Marland, G. (2013). Ranking of the World's Countries by 2009 Total

641   CO_2 Emissions from Fossil-fuel Burning, Cement Production, and Gas Flaring. Carbon

642   Dioxide Information Analysis Center. URL: cdiac.ornl.gov/trends/emis/top2013.tot.

643

644   Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S. W., Evan, S., McKeen, S. A., ... & Peischl,

645   J. (2013). Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale

646   inverse modeling technique: assessing anthropogenic emissions of CO, NO_x and CO_2 and

647   their impacts. Atmospheric Chemistry and Physics, 13(7), 3661-3677.

648

649   CARB, 2017. California Greenhouse Gas Emission Inventory - 2017 Edition, California Air

650   Resources Board. Available at http://www.arb.ca.gov/cc/inventory/data/data.htm

651

652   Chen, F., and J. Dudhia (2001), Coupling an advanced land surface hydrology model with the

653   Penn State NCAR MM5 modeling system. Part 1: Model implementation and sensitivity,

654   Mon. Weather Rev., 129, 569–585.

655

656   Chevallier, F., Maksyutov, S., Bousquet, P., Bréon, F.-M., Saito, R., Yoshida, Y., and Yokota,

657   T.: On the accuracy of the CO_2 surface fluxes to be estimated from the GOSAT observations,

658   Geophys. Res. Lett., 36, L19807, doi:10.1029/2009GL040108, 2009.

659

660   EDGAR. EDGAR Greenhouse Gas Emissions Inventory v4.2 FT2010, 2011. URL

661   http://edgar.jrc.ec.europa.eu/index.php.

662

663 Feng, S., Lauvaux, T., Newman, S., Rao, P., Ahmadov, R., Deng, A., ... & Gurney, K. R. (2016).

664   Los Angeles megacity: a high-resolution land–atmosphere modelling system for urban CO_2

665   emissions. Atmospheric Chemistry and Physics, 16(14), 9019-9045.

666

667 Fischer, M. L., Parazoo, N., Brophy, K., Cui, X., Jeong, S., Liu, J., ... & Graven, H. (2017).

668   Simulating estimation of California fossil fuel and biosphere carbon dioxide exchanges

669   combining in situ tower and satellite column observations. Journal of Geophysical Research:

670   Atmospheres, 122(6), 3653-3671.

671

672 Gerbig, C., Lin, J.C., Wofsy, S.C., Daube, B.C., Andrews, A.E., Stephens, B.B., Bakwin, P.S.

673   and Grainger, C.A., 2003. Toward constraining regional-scale fluxes of CO_2 with

674   atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-

675   oriented framework. Journal of Geophysical Research: Atmospheres, 108(D24).

676

677 Göckede, M., Michalak, A. M., Vickers, D., Turner, D. P., & Law, B. E. (2010). Atmospheric

678   inverse modeling to constrain regional-scale CO_2 budgets at high spatial and temporal

679   resolution. Journal of Geophysical Research: Atmospheres, 115(D15).

680

681 Graven, H. D., Guilderson, T. P., & Keeling, R. F. (2007). Methods for high-precision ^{14}C AMS

682   measurement of atmospheric CO_2 at LLNL. Radiocarbon, 49(2), 349-356.

683

30

684 Graven, H. D., Guilderson, T. P., & Keeling, R. F. (2012). Observations of radiocarbon in CO_2 at

685   seven global sampling sites in the Scripps flask network: Analysis of spatial gradients and

686   seasonal cycles. Journal of Geophysical Research: Atmospheres, 117(D2).

687

688 Graven, H., Fischer, M. L., Lueker, T., Jeong, S., Guilderson, T. P., Keeling, R. F., ... &

689   Frankenberg, C. (2018). Assessing fossil fuel CO_2 emissions in California using atmospheric

690   observations and models. Environmental Research Letters, 13(6), 065007.

691

692 Grell, G. A., Dudhia, J., & Stauffer, D. R. (1994). A description of the fifth-generation Penn

693   State/NCAR Mesoscale Model (MM5).

694

695 Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., & de la

696   Rue du Can, S. (2009). High resolution fossil fuel combustion CO_2 emission fluxes for the

697   United States. Environmental science & technology, 43(14), 5535-5541.

698

699 Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., ... & Fung, I.

700   Y. (2003). TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and

701   sensitivity to transport and prior flux information. Tellus B: Chemical and Physical

702   Meteorology, 55(2), 555-579.

703

704 Hogue, S., Marland, E., Andres, R. J., Marland, G., & Woodard, D. (2016). Uncertainty in

705   gridded CO_2 emissions estimates. Earth's Future, 4(5), 225-239.

706

31

707   Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit

708   treatment of entrainment processes. Monthly weather review, 134(9), 2318-2341.

709

710   Hungershoefer, K., Breon, F.-M., Peylin, P., Chevallier, F., Rayner, P., Klonecki, A., Houweling,

711   S., and Marshall, J.: Evaluation of various observing systems for the global monitoring of

712   CO_2 surface fluxes, Atmos. Chem. Phys., 10, 10503–10520, doi:10.5194/acp-10-10503-

713   2010, 2010.

714

715   IPCC. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and II to

716   the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [Core

717   Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. Technical report, 2014.

718

719   Jacob, D. J. (2007). Lectures on inverse modeling. Harvard University.

720

721   Jeong, S., Hsu, Y. K., Andrews, A. E., Bianco, L., Vaca, P., Wilczak, J. M., & Fischer, M. L.

722   (2013). A multitower measurement network estimate of California's methane emissions.

723   Journal of Geophysical Research: Atmospheres, 118(19), 11-339.

724

725   Jeong, S., Newman, S., Zhang, J., Andrews, A. E., Bianco, L., Bagley, J., ... & LaFranchi, B. W.

726   (2016). Estimating methane emissions in California's urban and rural regions using

727   multitower observations. Journal of Geophysical Research: Atmospheres, 121(21).

728

729 Jones, A., Thomson, D., Hort, M., & Devenish, B. (2007). The UK Met Office's next-generation

730   atmospheric dispersion model, NAME III. In Air pollution modeling and its application XVII

731   (pp. 580-589). Springer, Boston, MA.

732

733 Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., & Meijer,

734   H. A. (2005). Atmospheric CO_2 and $^{13}CO_2$ exchange with the terrestrial biosphere and oceans

735   from 1978 to 2000: observations and carbon cycle implications. In A history of atmospheric

736   CO2 and its effects on plants, animals, and ecosystems (pp. 83-113). Springer, New York,

737   NY.

738

739 Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., ... &

740   Song, Y. (2016). High-resolution atmospheric inversion of urban CO_2 emissions during the

741   dormant season of the Indianapolis Flux Experiment (INFLUX). Journal of Geophysical

742   Research: Atmospheres, 121(10), 5213-5236.

743

744 Levin, I., Kromer, B., Schmidt, M., & Sartorius, H. (2003). A novel approach for independent

745   budgeting of fossil fuel CO2 over Europe by $^{14}CO_2$ observations. Geophysical Research

746   Letters, 30(23).

747

748 Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., & Derwent, R. G. (2011).

749   Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion

750   modeling approach. Journal of Geophysical Research: Atmospheres, 116(D2).

751

752 Marland, G., Brenkert, A., & Olivier, J. (1999). CO2 from fossil fuel burning: a comparison of

753   ORNL and EDGAR estimates of national emissions. Environmental Science & Policy, 2(3),

754   265-273.

755

756 Nakanishi, M., & Niino, H. (2004). An improved Mellor–Yamada level-3 model with

757   condensation physics: Its design and verification. Boundary-Layer Meteorology, 112(1), 1-

758   31.

759

760 Nakanishi, M., & Niino, H. (2006). An improved Mellor–Yamada level-3 model: Its numerical

761   stability and application to a regional prediction of advection fog. Boundary-Layer

762   Meteorology, 119(2), 397-407.

763

764 Nassar, R., Sioris, C. E., Jones, D. B. A., and McConnell, J. C.: Satellite observations of CO_2

765   from a highly elliptical orbit for studies of the Arctic and boreal carbon cycle, J. Geophys.

766   Res.- Atmos., 119, 2654–2673, doi:10.1002/2013JD020337, 2014.

767

768 Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., & Deng, F.

769   (2013). Improving the temporal and spatial distribution of CO_2 emissions from global fossil

770   fuel emission data sets. Journal of Geophysical Research: Atmospheres, 118(2), 917-933.

771

772 Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., & Freitas, S.

773   (2010). Coupled weather research and forecasting–stochastic time-inverted lagrangian

774   transport (WRF–STILT) model. Meteorology and Atmospheric Physics, 107(1-2), 51-64.

34

775

776   Newman, S., Jeong, S., Fischer, M. L., Xu, X., Haman, C. L., Lefer, B., ... & Peischl, J. (2013).

777   Diurnal tracking of anthropogenic CO_2 emissions in the Los Angeles basin megacity during

778   spring 2010. Atmospheric Chemistry and Physics, 13(8), 4359-4372.

779

780   Oda, T., & Maksyutov, S. (2011). A very high-resolution (1 km × 1 km) global fossil fuel CO_2

781   emission inventory derived using a point source database and satellite observations of

782   nighttime lights. Atmospheric Chemistry and Physics, 11(2), 543-556.

783

784   Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., ... & Delage, F.

785   (2011). Importance of fossil fuel emission uncertainties over Europe for CO_2 modeling:

786   model intercomparison. Atmospheric chemistry and physics, 11(13), 6607-6622.

787

788   Peylin P., Law RM, Gurney KR, Chevallier F, Jacobson AR, Maki T, Niwa Y, Patra PK, Peters

789   W, Rayner PJ, Rödenbeck C. Global atmospheric carbon budget: results from an ensemble of

790   atmospheric CO2 inversions. Biogeosciences. 2013 Oct 24;10:6699-720.

791

792   Rodgers, C. D. (2000). Inverse methods for atmospheric sounding: theory and practice. World

793   Scientific. ISBN 978-981-02-2740-1. doi: 10.1142/3171. URL

794   http://www.worldscientific.com/worldscibooks/10.1142/3171.

795

796   Seibert, P., Beyrich, F., Gryning, S. E., Joffre, S., Rasmussen, A., & Tercier, P. (2000). Review

797   and intercomparison of operational methods for the determination of the mixing height.

798   Atmospheric environment, 34(7), 1001-1027.

799

800   Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., ... &

801   Saripalli, S. (2011). Assessment of fossil fuel carbon dioxide and other anthropogenic trace

802   gas emissions from airborne measurements over Sacramento, California in spring 2009.

803   Atmospheric Chemistry and Physics, 11(2), 705-721.

804

805   Turnbull, J. C., Sweeney, C., Karion, A., Newberger, T., Lehman, S. J., Tans, P. P., ... &

806   Cambaliza, M. O. (2015). Toward quantification and source sector identification of fossil

807   fuel CO2 emissions from an urban area: Results from the INFLUX experiment. Journal of

808   Geophysical Research: Atmospheres, 120(1), 292-312.

809

810   Yver, C. E., Graven, H. D., Lucas, D. D., Cameron-Smith, P. J., Keeling, R. F., & Weiss, R. F.

811   (2013). Evaluating transport in the WRF model along the California coast. Atmospheric

812   Chemistry and Physics, 13(4), 1837-1852.

Air Basin	Name	Code	Vulcan (TgC/yr)	EDGAR (TgC/yr)	SD Prior Unc. 1-σ (%)	Vulcan - EDGAR (TgC/yr)
1	North Coast	1.NC	1.0	1.6	59	-0.6
2	Northeast Plateau	2.NP	0.4	1.3	96	-1.0
3	Sacramento Valley	3.SV	6.8	7.4	8	-0.7
4	Mountain Counties	4.MC	2.2	2.0	51	0.1
5	Lake County	5.LC	0.1	0.2	65	-0.2
6	Lake Tahoe	6.LT	0.1	0.1	42	0
7	Great Basin Valleys	7.GBV	0.2	0.6	100	-0.4
8	San Joaquin Valley	8.SJV	8.6	20.2	35	-11.6
9	North Central Coast	9.NCC	6.0	2.2	71	3.8
10	Mojave Desert	10.MD	6.1	4.3	17	1.8
11	South Central Coast	11.SCC	4.4	3.4	21	1.0
12	Salton Sea	12.SS	1.4	1.7	55	-0.3
13	San Francisco Bay	13.SFB	16.4	17.5	22	-1.2
14	South Coast	14.SC	26.9	35.5	12	-8.6
15	San Diego	15.SD	6.6	6.5	10	0.1
Total California			**89.6**	**104.7**	**8**	**-17.8**

Table 1: The 15 air basins of California with respective emissions as estimated by Vulcan and EDGAR. Also shown are the SD prior uncertainty estimate (Fischer et al., 2017), and difference in magnitude between Vulcan and EDGAR for each air basin. Air basin numbers correspond to those marked in Figure 1.
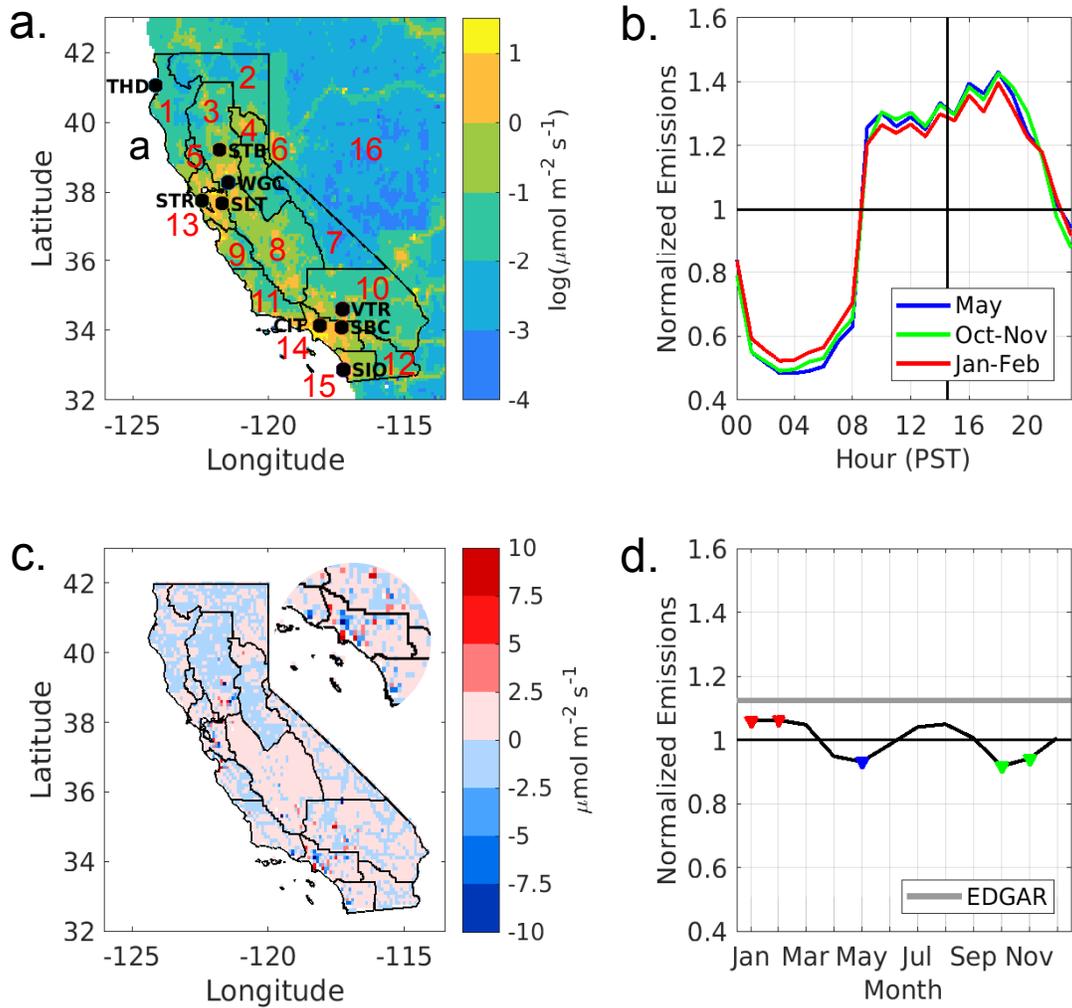
Figure 1: **a.** The location of the 9 tower sites in the observation network (marked with black circles): Trinidad Head (THD), Sutter Buttes (STB), Walnut Grove (WGC), Sutro (STR), Sandia-Livermore (LVR), Victorville (VTR), San Bernardino (SBC), Caltech (CIT) and Scripps Institute of Oceanography (SIO). The 15 air basins are marked out with black lines with region 16 representing emission from outside California within the model domain. Underlayed is a map of annual mean ffCO$_2$ emissions from the Vulcan v2.2 emission map within the United States and EDGAR v4.2 (FT2010) for emission from outside the U.S. **b.** Vulcan diurnal emissions normalized to campaign averaged emissions for the 3 campaigns, **c.** Scaled EDGAR subtracted from Vulcan emissions map, where EDGAR has been scaled to have the same air basin total emissions. The inset shows an enlarged view of southwestern California. **d.** Average monthly emissions normalized to Vulcan annual emissions. Shown in both **b** and **d** is EDGAR annual invariant emissions (grey).

| Transport Model | Meteorology | Domain | Model Resolution | | | References |
			Horizontal	Vertical (nLevels / Max Height)	Temporal	
WS-LBL	WRF (v3.5.1)	North America	1km, 4km, 12 km ,36km	50 / 16 km	1 hour	(Lin, 2003; Nehrkorn et al., 2010);
WS-CTL	WRF (v2.1.2)	North America	0.1°, 1°	29 / 25 km	1 hour	(Carbon Tracker, 2017)
UM-NAME	Unified Model	Global	17km, 25km	59 / 29 km	3 hours	(Ryall et al., 1998)

Table 2: Comparison of the three atmospheric transport models used in this study.

3
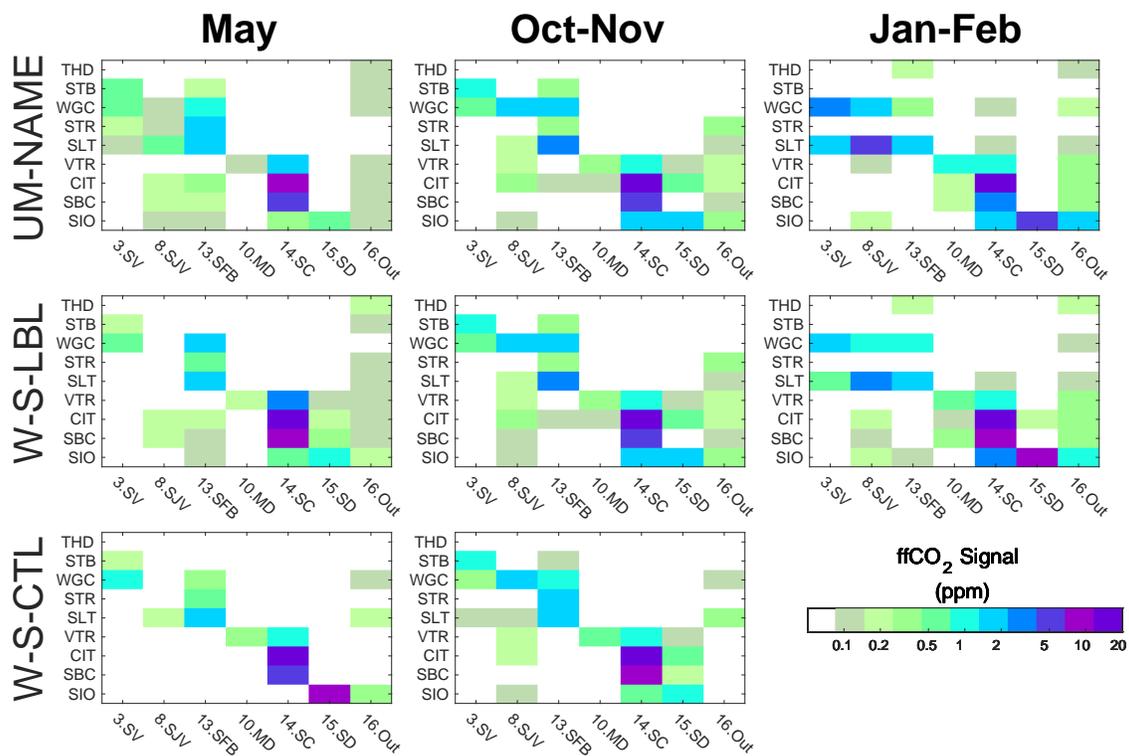
Figure 2: The average ffCO₂ signal (ppm) simulated by each atmospheric transport model as a result of emissions from the 6 largest emitting air basins and one outside California region at each observation site over the three measurement campaigns. Signals were simulated based on the EDGAR emission map.
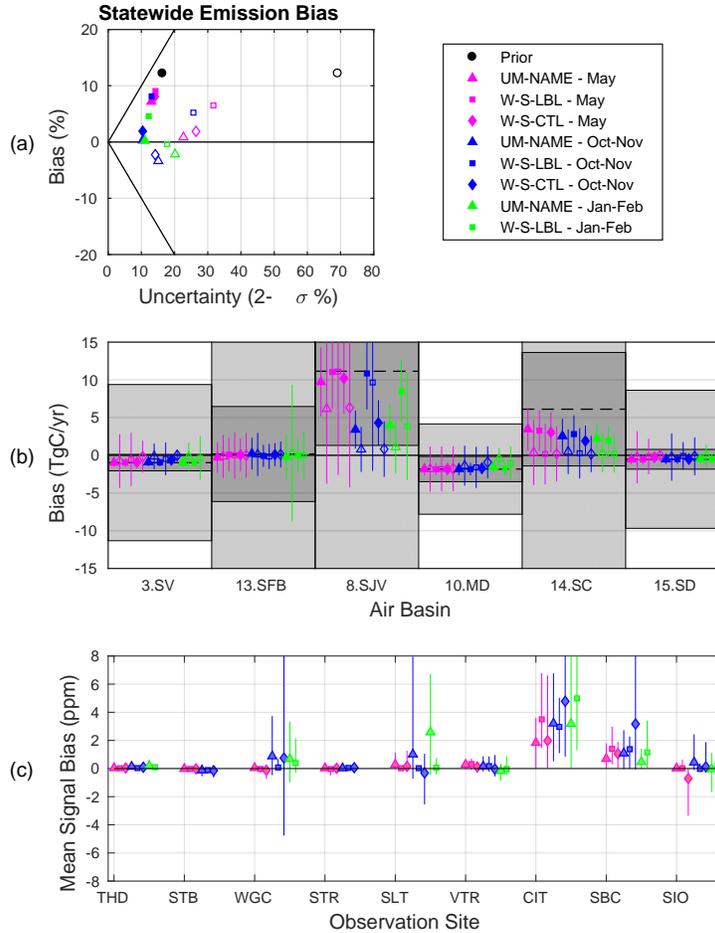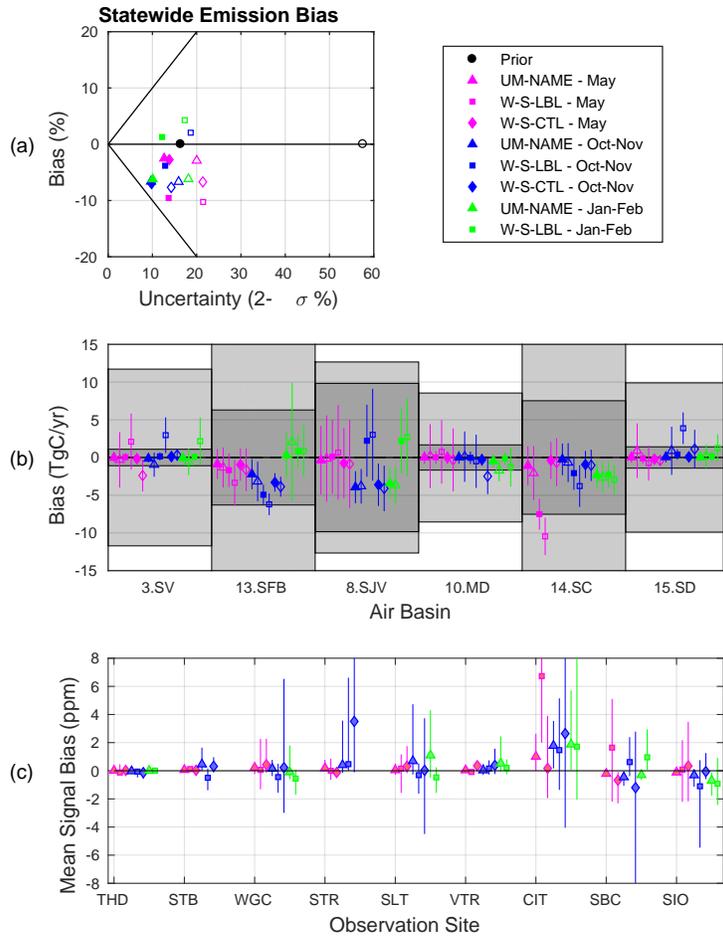
Figure 3: (a) Statewide and (b) individual air basin inversion results for an error in the magnitude of prior emissions. Prior emissions are given by EDGAR and true emissions are given by EDGAR scaled to Vulcan total in each air basin. Air basin results are shown for Sacramento Valley (3.SV), San Francisco Bay (13.SFB), San Joaquin Valley (8.SJV), Mojave Desert (10.MD), South Coast (14.SC) and San Diego (15.SD). Prior results are presented by black markers and posterior results are represented by colored markers. Filled markers show results using SD prior uncertainty and empty markers show results using 70% prior uncertainty. The prior bias in each air basin is given by the dashed lines in (b) with SD prior uncertainty (dark grey) and 70% prior uncertainty (light grey). Prior and posterior uncertainties are expressed as 2-σ. The bottom plot (c) shows the mean signal error in simulated average ffCO$_2$ concentration. Mean signal error is calculated by subtracting the average true signal from the average prior signal. Error lines are drawn between the maximum and minimum signal bias per campaign.

5

Figure 4: (a) Statewide and (b) individual air basin inversion results for an error in the spatial distribution of prior emissions. Prior emissions are given by EDGAR scaled to Vulcan emissions totals in each air basin and true emissions are given by Vulcan. The bottom plot (c) shows the mean signal error in simulated average ffCO$_2$ concentration.
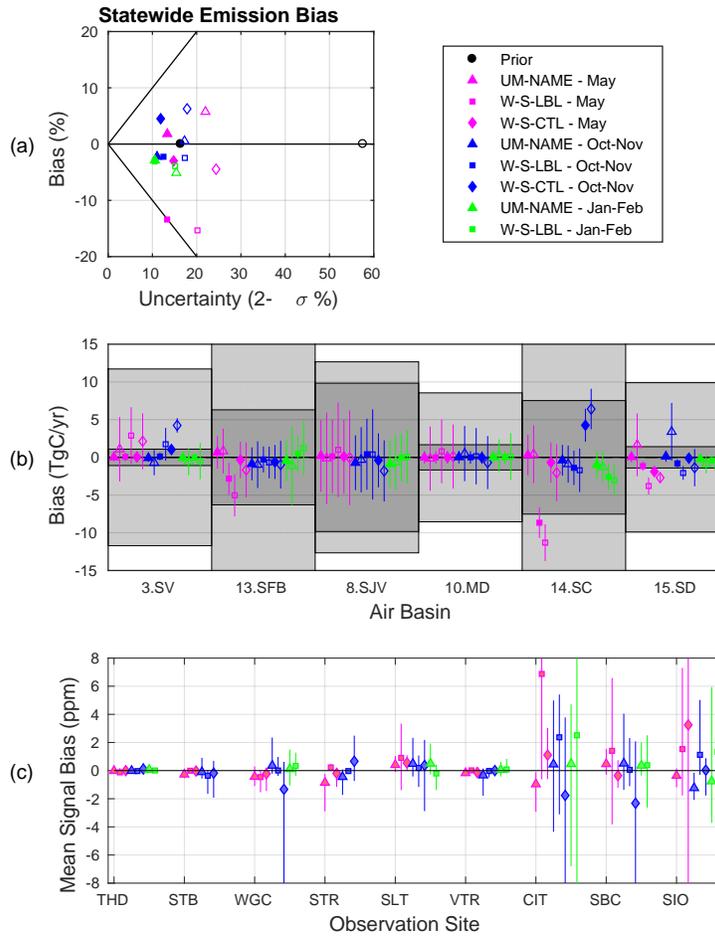
Figure 5: (a) Statewide and (b) individual air basin inversion results for an error in the temporal distribution of prior emissions. Prior emissions are given by temporally varying Vulcan and true emissions are given by annually averaged Vulcan. Prior emissions were scaled to be the equal in magnitude to annually averaged Vulcan emissions. The bottom plot (c) shows the mean signal error in simulated average ffCO$_2$ concentration.
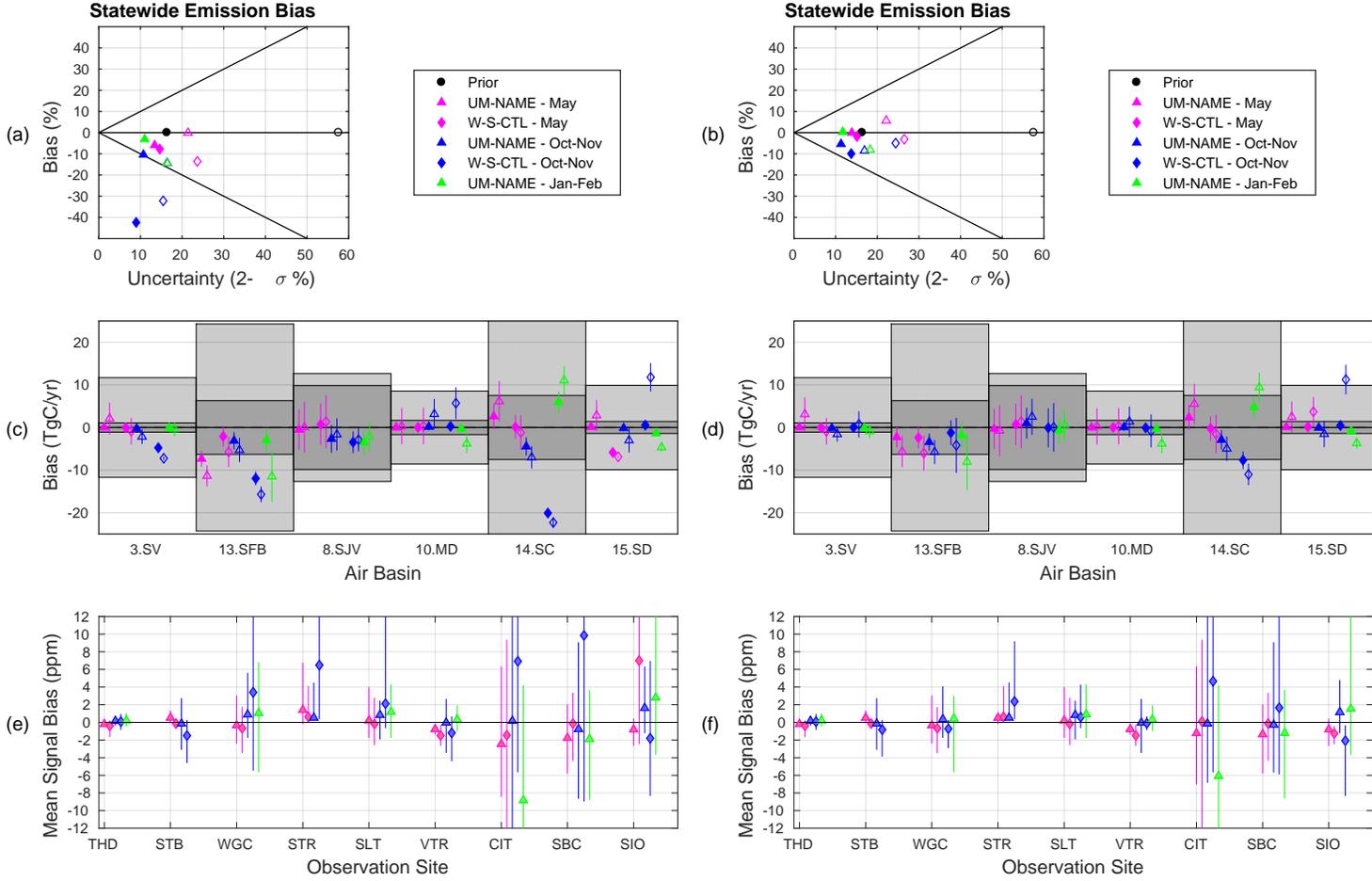
Figure 6: Inversion results for the experiment where the atmospheric transport in the prior simulation uses WS-CTL or UM-NAME but the atmospheric transport in the true simulation uses WS-LBL. Posterior statewide emissions (a, b), individual air basin emissions (c, d), and percentage error in simulated average ffCO$_2$ concentration (e, f) are shown with no outlier removal (first column) and outliers removed (second column). Prior and true emissions are given by annually averaged Vulcan.
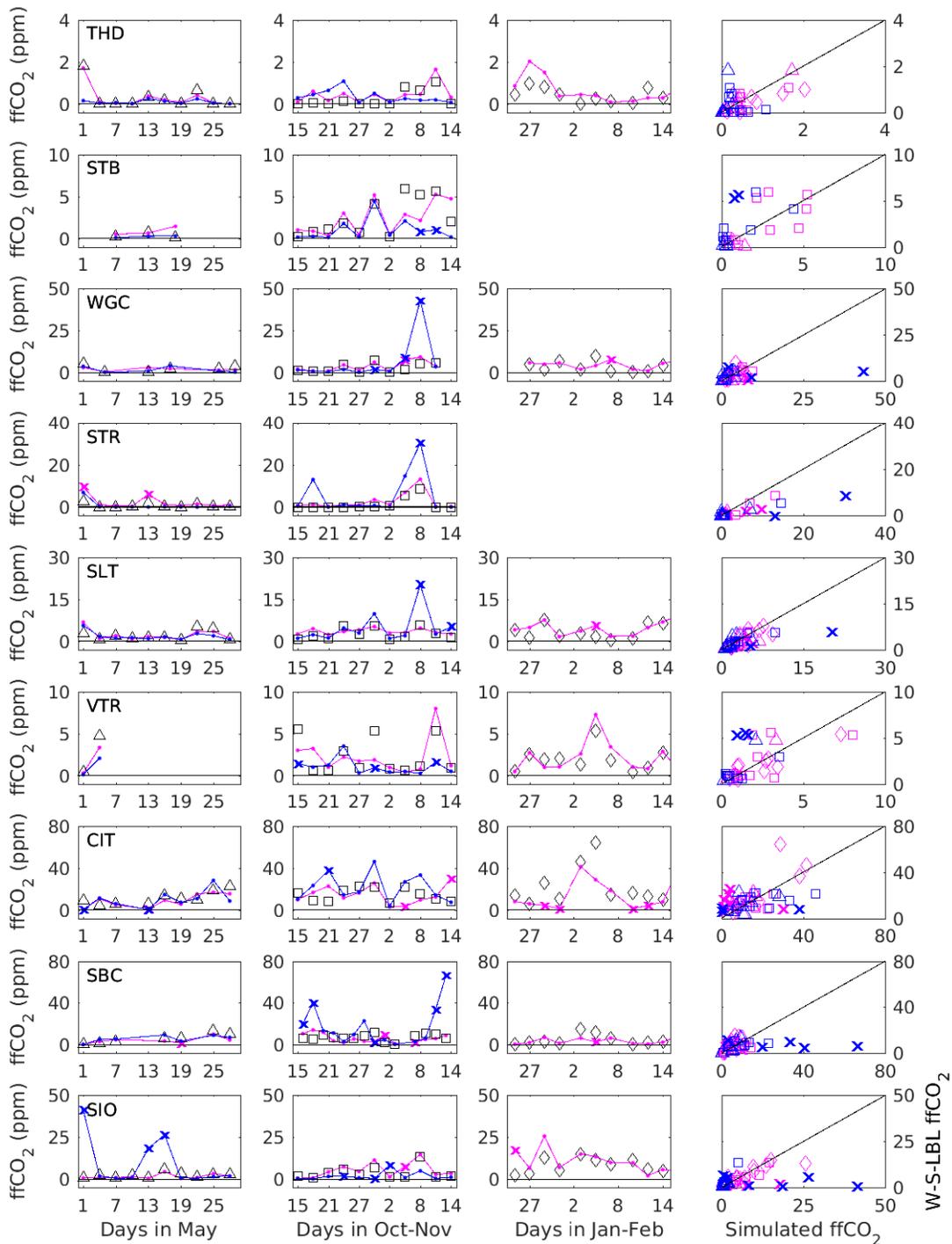
Figure 7: All simulated ffCO$_2$ from May (first column), October-November (second column), and January-February (third column). Simulated ffCO$_2$ using W-S-LBL are shown in black markers (triangles for May, squares for Oct-Nov and diamonds for Jan-Feb) whilst prior W-S-CTL signals are shown in blue and UM-NAME signals in magenta. All simulated signals are generated using the Vulcan gridded emissions map. The fourth column shows true vs prior ffCO$_2$ signals, with colors corresponding to models and markers corresponding to campaigns. Outliers omitted from the standard inversion are shown by an x.

9

Supplementary
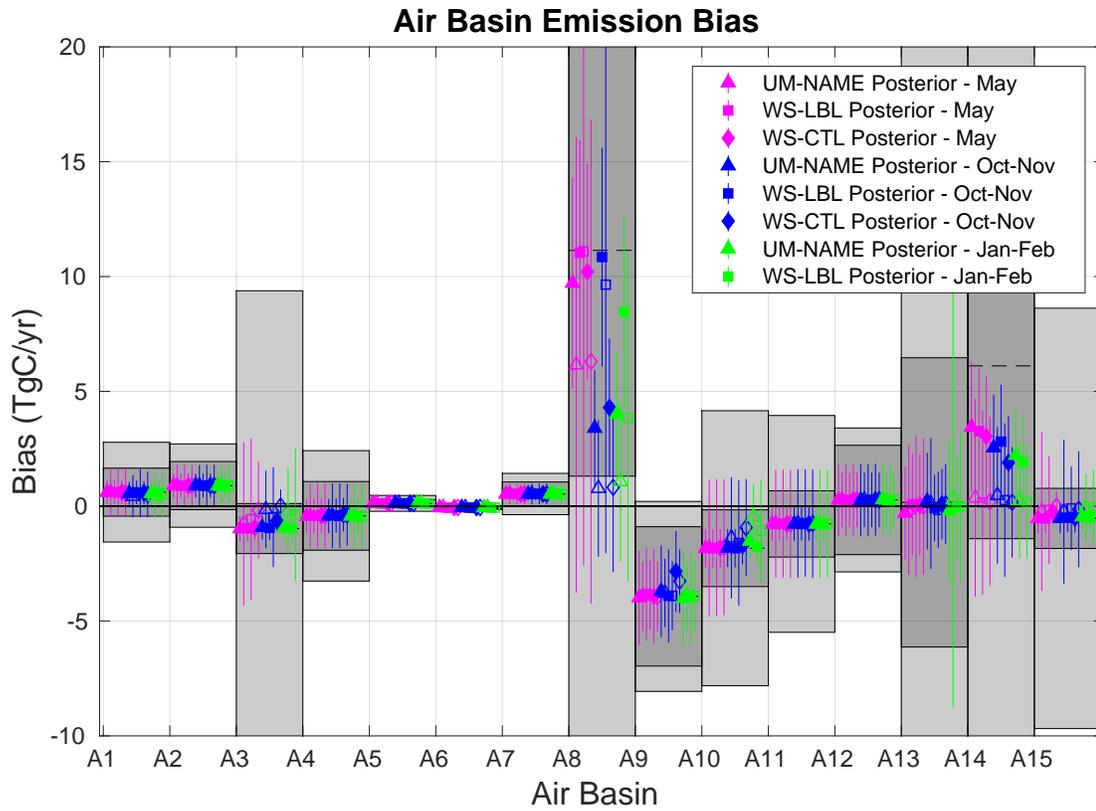
Figure 1: Posterior air basin emissions for an error in the magnitude of emission. Filled markers show posterior results using SD prior uncertainty and clear markers represent 70% prior uncertainty. The prior bias in each air basin is given by the dashed lines with SD prior uncertainty (dark grey) and 70% prior uncertainty (light grey). Prior and posterior uncertainties are expressed as 2-σ.
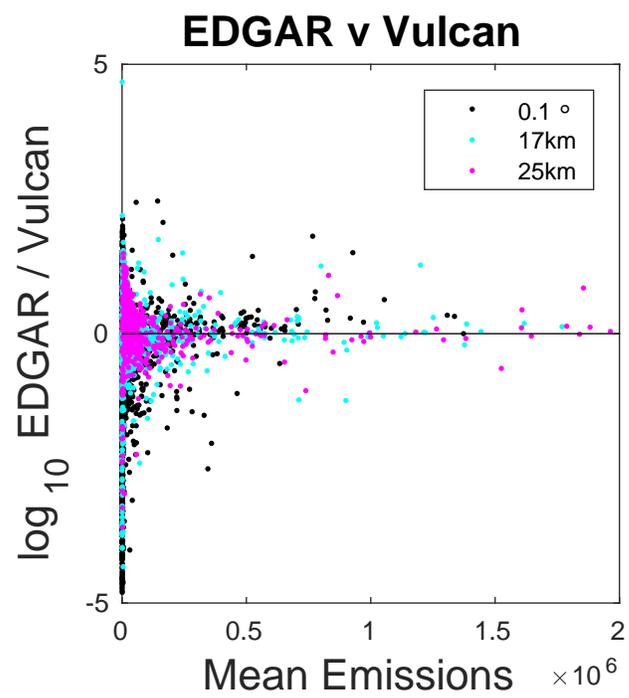
Figure 2: EDGAR grid cells compared to Vulcan (regridded to EDGAR native 0.1° resolution). Mean emissions are in units of gCO$_2$ m^2 yr^{-1}.
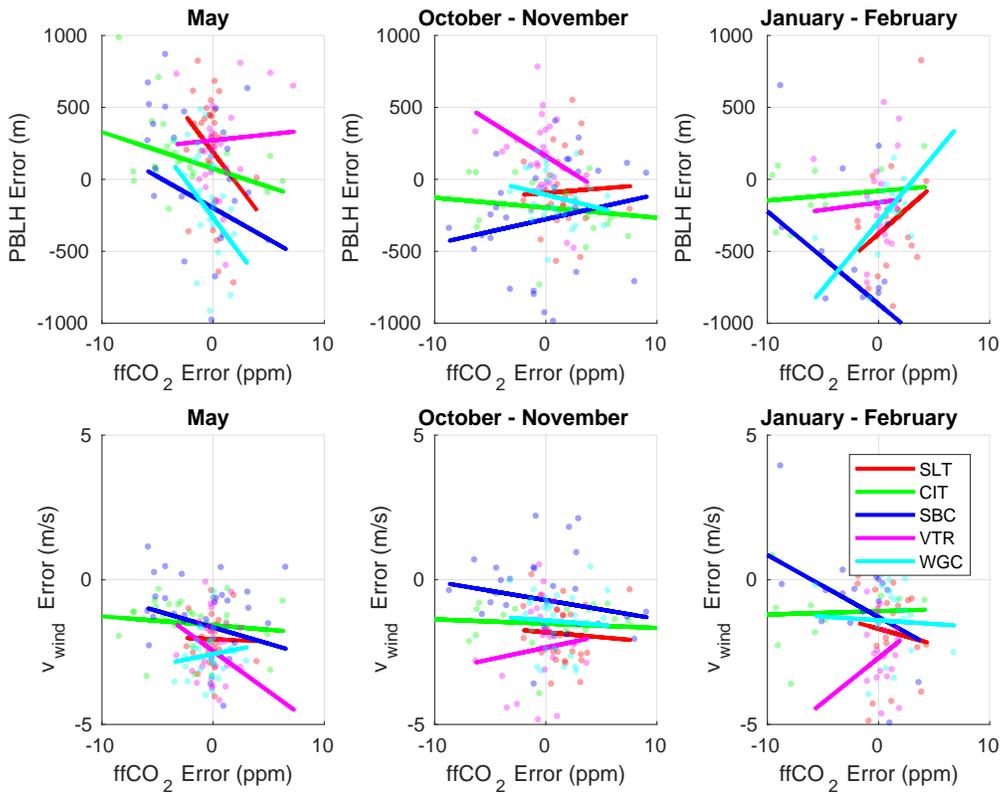
Figure 3: PBLH error versus ffCO$_2$ error (top) and wind speed versus ffCO$_2$ error (bottom). Error in PBLH/wind speed was calculated by subtracting UM from WRF estimates. ffCO$_2$ signal error was calculated by subtracting UM-NAME from WS-LBL signals.
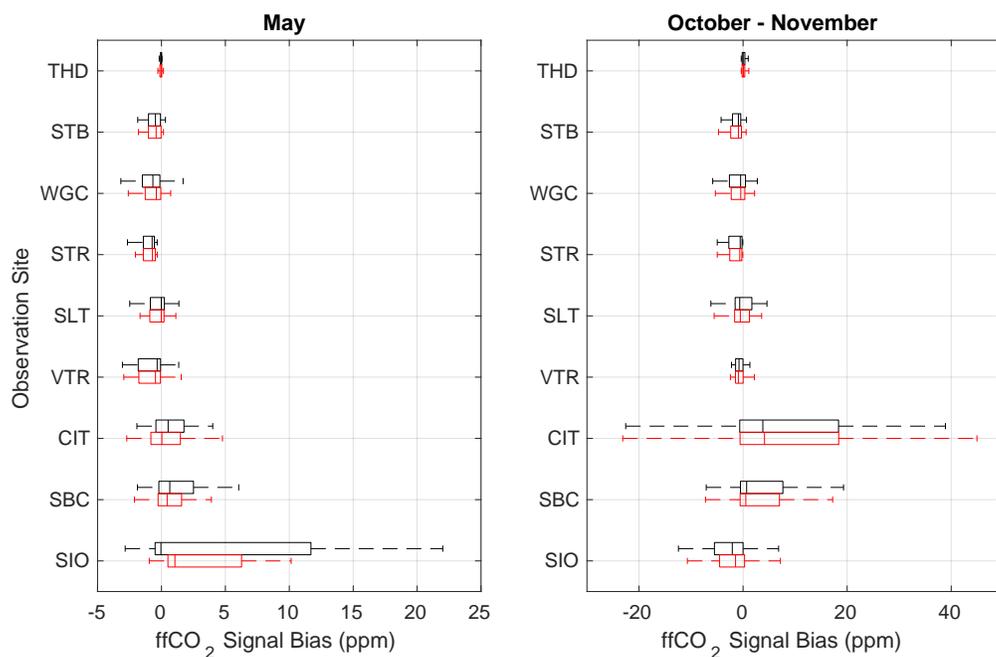
Figure 4: ffCO$_2$ signal bias for aggregation error (black) and no aggregation error (red) in modelled atmospheric transport. For each box the central mark indicates the median, and the left and right edges indicate the 25th and 75th percentiles, respectively. Dashed lines extend to the most extreme data points not considered outliers. Error was calculated by subtracting WS-CTL signals (generated using native 0.1° resolution and UM-NAME resolution footprints respectively) from UM-NAME signals.
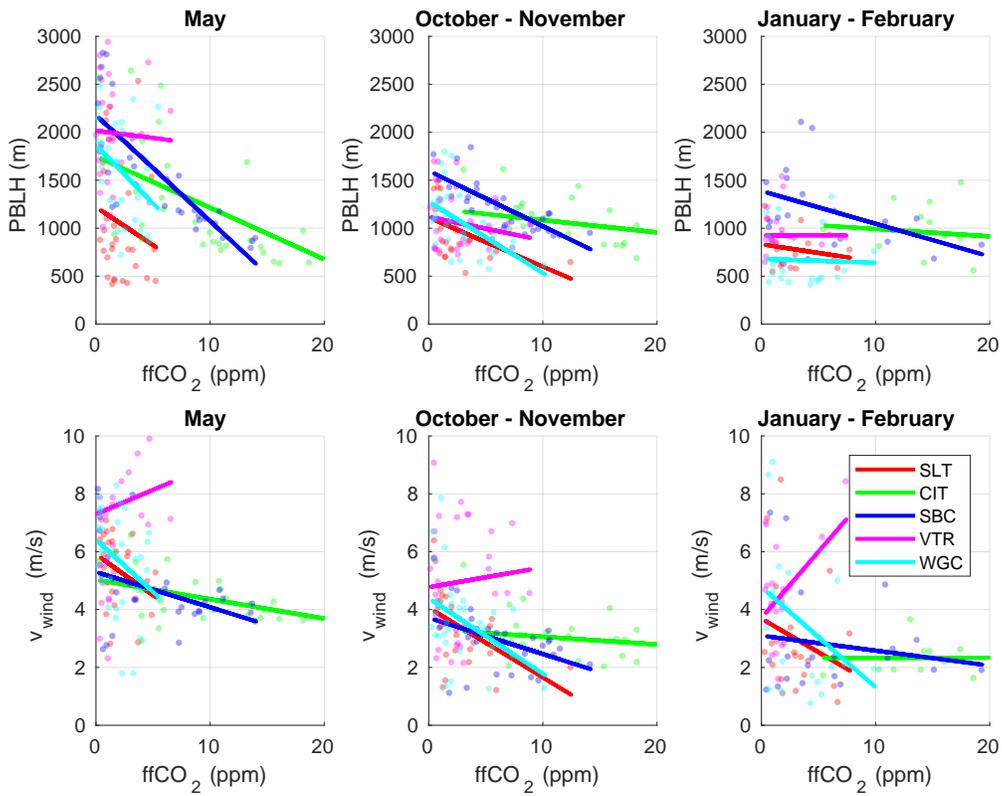
Figure 5: PBLH versus ffCO$_2$ (top) and wind speed versus ffCO$_2$ (bottom) using MYNN2 PBLH, Noah/LSU wind speed, and WS-LBL ffCO$_2$ signal estimates.