Title
Estimating methane emissions in California’s urban and rural regions using multi-tower observations:

Authors
Jeong, Seongeun
Newman, Sally
Zhang, Jingsong
et al.

Publication Date
2016-07-01

LBNL-1005846

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Estimating methane emissions in California's urban and rural regions using multi-tower observations

Seongeun Jeong, Sally Newman, Jingsong Zhang, Arlyn E. Andrews, Laura Bianco Justin Bagley, Xinguang Cui, Heather Graven, Jooil Kim, Peter Salameh, Brian W. LaFranchi, Chad Priest, Mixtli Campos-Pineda, Elena Novakovskaia, Christopher D. Sloop, Hope A. Michelsen, Ray P. Bambha, Ray F. Weiss, Ralph Keeling, and Marc L. Fischer

Energy Technologies Area

July 2016

Pre-print version submitted to Journal of Geophysical Research (JGR); pending approval.

This work was supported by the U.S. Department of Energy under Lawrence Berkeley National Laboratory Contract No. DE-AC02-05CH11231.

Estimating methane emissions in California's urban and rural regions using multi-tower observations

Seongeun Jeong[1], Sally Newman[2], Jingsong Zhang[3], Arlyn E. Andrews[4], Laura Bianco[4,5] Justin Bagley[1], Xinguang Cui[1], Heather Graven[6], Jooil Kim[7], Peter Salameh[7], Brian W. LaFranchi[8], Chad Priest[3], Mixtli Campos-Pineda[3], Elena Novakovskaia[9], Christopher D. Sloop[9], Hope A. Michelsen[8], Ray P. Bambha[8], Ray F. Weiss[7], Ralph Keeling[7], Marc L. Fischer[1]

1. Lawrence Berkeley National Lab, Berkeley, CA
2. California Institute of Technology, Pasadena, CA
3. University of California, Riverside, CA
4. ESRL, NOAA, Boulder, CO
5. CIRES, Univ. of Colorado, Boulder, CO
6. Imperial College London, London, UK
7. Scripps Institution of Oceanography, UC San Diego, La Jolla, CA
8. Sandia National Labs, Livermore, CA
9. Earth Networks, Germantown, MD

Corresponding author:
Seongeun Jeong
MS 90R2002
1 Cyclotron Rd.
Berkeley, CA 94720 USA
phone: 510-486-7216
email: sjeong@lbl.gov

Key Points

- Multi-site observations constrain both urban and rural CH_4 emissions

- California total emissions are likely 1.2 -1.8 times state inventory

- More efforts are needed to constrain emissions by both sector and region

Abstract

We present an analysis of methane (CH_4) emissions using atmospheric observations from thirteen sites in California during June 2013 – May 2014. A hierarchical Bayesian inversion method is used to estimate CH_4 emissions for spatial regions (0.3° pixels for major regions) by comparing measured CH_4 mixing ratios with transport model (WRF-STILT) predictions based on seasonally varying California-specific CH_4 prior emission models. The transport model is assessed using a combination of meteorological and carbon monoxide (CO) measurements coupled with the gridded California Air Resources Board (CARB) carbon monoxide (CO) emission inventory. Hierarchical Bayesian inversion suggests that state annual anthropogenic CH_4 emissions are 2.42 ± 0.49 Tg CH_4/yr (at 95% confidence, including transport bias uncertainty), higher (1.2 - 1.8 times) than the CARB current inventory (1.64 Tg CH_4/yr in 2013). We note that the estimated CH_4 emissions drop to 1.0 - 1.6 times the CARB inventory if we correct for the 10% median CH_4 emissions assuming the bias in CO analysis is applicable to CH_4. The CH_4 emissions from the Central Valley and urban regions (San Francisco Bay and South Coast Air Basins) account for ~58% and 26% of the total posterior emissions, respectively. This study suggests that the livestock sector is likely the major contributor to the state total CH_4 emissions, in agreement with CARB's inventory. Attribution to source sectors for sub-regions of California using additional trace gas species would further improve the quantification of California's CH_4 emissions and mitigation efforts towards the California Global Warming Solutions Act of 2006 (AB-32).

Keywords: methane, greenhouse gas, emission inventory, atmospheric transport, inverse model

Index Terms: 0365, 0345, 0368

1. Introduction

California has committed to an ambitious plan to reduce greenhouse gas (GHG) emissions to 1990 levels by 2020 through Assembly Bill 32 (AB-32), which requires accurate accounting of CH_4 emissions for effective mitigation planning and verification of future emission reductions. The state official GHG inventory reports that California currently emits a total of approximately 459.3 Tg CO_2 (1 Tg = 10^{12} g) equivalent GHGs each year [California Air Resources Board (CARB), 2015]. The CARB GHG inventory is produced in support of AB-32, thus only includes anthropogenic emission sources. Among the reported GHGs, ~9% of the total GHG emissions are attributed to methane (CH_4), which is the second largest contributor to climate forcing emissions in California behind carbon dioxide (CO_2) [CARB, 2015]. Moreover, as shown in previous studies (e.g., Jeong et al. [2013, 2014]) CH_4 emissions in California are relatively uncertain compared to those of CO_2 due to lack of activity data and incomplete understanding of emission processes, and top-down studies can be complicated by California's diverse emission sources, complex topography and weather patterns.

Several recent studies have estimated CH_4 emissions in different regions of California using measurements from ground towers, aircrafts, and satellites. At the regional scale, Zhao et al. [2009] and Jeong et al. [2012a, 2013] estimated CH_4 emissions using towers in the Central Valley. In particular, Jeong et al. [2013] conducted the first multi-site analysis of CH_4 emissions in California based on measurements from five ground sites and across seasons (ten months during 2010 - 2011), and estimated a state total of 2.03 – 2.71 Tg CH_4/yr (at 68% confidence). Wecht et al. [2014] used airborne measurements during a short-period campaign (May – June

2010) and estimated a total of 2.65 – 3.07 Tg CH_4/yr (at 68% confidence) based on a different prior emission model that resulted in a different source apportionment from that of Jeong et al. [2013], attributing significantly higher emissions to landfill and wastewater.

At the sub-regional scale, most studies focused on the urban regions of southern California [e.g., Wunch et al., 2009; Hsu et al., 2010; Wennberg et al., 2012; Peischl et al., 2013]. Although the urban studies relied on different analysis methods (e.g., ratio of CH_4 to CO (carbon monoxide)) and measured data from different years, the focus region for each study generally covered the South Coast Air Basin (SoCAB) of California. For SoCAB, the estimated CH_4 emissions ranged from 280 to 700 Gg CH_4/yr (based on the reported uncertainty estimates, 1 Gg = 10^9 g). In another study, Jeong et al. [2014] estimated statewide CH_4 emissions from petroleum production and the natural gas system, taking a unique approach of combining a bottom-up inventory with results from a field campaign.

Here we expand on previous work by Jeong et al. [2012a, 2013, 2014] to quantify both urban and rural CH_4 emissions from California, presenting the first analysis of full annual CH_4 emissions from California using atmospheric observations from 13 tower sites covering all major CH_4-emitting regions of California. In particular, this study uses the hierarchical Bayesian approach introduced by Ganesan et al. [2014] for the purpose of GHG emission quantification. In this study we illustrate how uncertainty in the inversion can be treated by a combination of our best a priori knowledge of error sources (e.g., transport error) and statistical inference, and how ground-based multi-tower measurements can be effectively used to constrain regional emissions. In Section 2, we describe the methods we employed, including atmospheric measurements, a priori CH_4 emissions, transport modeling, and the hierarchical Bayesian inverse method. Section 3 presents results, including the inferred CH_4 emissions from California for different regions and sources. Section 4 further discusses the results and presents conclusions for CH_4 emissions in California.

2. Data and Methods

2.1. CH_4 Measurements and Background

CH_4 measurements were made at the collaborative 13-site GHG network across California during June 2013 – May 2014. The information of sites and data availability is summarized in Table 1 (see Figure 1 for site locations). Detailed information regarding measurement methods for the Central Valley sites are summarized in Jeong et al. [2012a, 2013] and Andrews et al. [2014]. Here, we briefly describe measurements as a component of the inverse modeling framework. All sites are operated with temperature and pressure-controlled cavity ring-down CH_4 gas analyzers (Picarro Inc.), permeation-tube gas sample driers, and periodic calibrations using either primary NOAA (National Oceanic and Atmospheric Administration) CH_4 gas standards or secondary gas standards. For this study, we added four new sites in southern California: CIT (Caltech), SBC (San Bernardino), SIO (Scripps Institution of Oceanography) and VTR (Victorville), and two sites for the San Francisco Bay Area (SFBA): LVR (Livermore) and STR (Sutro tower). All new sites except STR had similar instrumentation to existing sites, while STR employed daily flask samples collected for approximately 2 minutes near 1400 Local Standard Time (LST) for subsequent analysis at NOAA Earth System Research Laboratory. Measurements at THD (Trinidad Head) were made by a flame ionization gas chromatography (FIGC) system as part of the Advanced Global Atmospheric Gases Experiment (AGAGE) network [Prinn et al., 2000]. The Tohoku University calibration scale used by AGAGE is indistinguishable from the NOAA04 calibration scale used for our Picarro measurements, with a relative scale factor of 1.0003 [Hall et al., 2014]. Thus, no corrections for scale differences were applied. In addition, we assume that the isotopic effect in transferring the NOAA standards that are calibrated by FIGC measuring all CH_4 isotopologues to the Picarro instrument (measuring only the predominant CH_4 isotopologue) is negligible. For continuous measurement sites, calibrated data were averaged to hourly intervals and then 3-hourly intervals for inversions following the procedure in Jeong et al. [2012a, 2013]. All sites are expected to provide measurement precision that is smaller than the CH_4 synoptic variations typically observed in the ambient air, and with absolute accuracy sufficient to provide negligible bias in estimating the scaling relationship between observed and predicted CH_4 signals.

Following previous work (e.g., Jeong et al. [2013]), we selected measurements that coincided with periods when the atmospheric boundary layer was well-mixed. For the Walnut Grove tower (WGC) we explicitly evaluated atmospheric mixing using measured vertical CH_4 profiles. As in Jeong et al. [2012a, 2013], WGC data from 91 m were selected in the time window between 1200 and 1700 LST, subject to the requirement that the CH_4 mixing ratio difference ($C_{91} – C_{483}$) between 91 and 483 m fell within the range $–1$ sd < ($C_{91} – C_{483}$) < 3 sd, where sd is the standard deviation of the 91-483 m difference. This additional requirement retained approximately 80% of data in the 1200 – 1700 LST window. We selected all data in the afternoon time window (1200 – 1700 LST) for other sites without profile information.

The predicted CH_4 upstream boundary values were estimated using a similar method to the one used in Jeong et al. [2012b, 2013]. The details for estimating the boundary values are described in Jeong et al. [2013] and only a summary is provided here. CH_4 boundary values were estimated using data from the Pacific coast aircraft network CH_4 profiles (http://www.esrl.noaa.gov/gmd/ccgg/aircraft/) and remote Pacific marine boundary layer sampling sites (http://www.esrl.noaa.gov/gmd/ccgg/flask.html) within the NOAA Earth System Research Laboratory (ESRL) Cooperative Air Sampling Network. The data were smoothed and interpolated to create a three-dimensional (3-D) curtain, varying with latitude, height and time. To quantify the errors associated with the 3-D curtain, we fit a smooth curve through the data and computed the seasonal cycle of the root mean square of the residuals from the curve. Predicted background values were computed for each hourly footprint simulation by sampling the curtain at each of the 500 particle trajectory endpoints (near the domain boundary at 130°W) and calculating the average value.

2.2. A priori CH_4 Emission Model

This work used the California Greenhouse Gas Emission Measurements (CALGEM) project a priori CH_4 emission model (henceforth CALGEM model, available at calgem.lbl.gov) described by Jeong et al. [2012a, 2013, 2014] with some modifications. The CALGEM emission model provides emissions by sector at a high spatial resolution (0.1° × 0.1°) for California. The CALGEM model has seasonal components for wetlands and crop agriculture only, and these seasonal emissions are combined with non-seasonal emissions to construct monthly emission maps for inversions. The inversion approach using non-seasonal prior emissions is widely used (e.g., Zhao et al. [2009], Jeong et al. [2012a; 2012b; 2013], Wecht et al. [2014], Cui et al.

[2015]). In particular, Jeong et al. [2012a; 2012b; 2013] showed non-seasonal priors can provide information on seasonality in the posterior emission.

In this study, the CALGEM prior emission distributions are scaled to match 2012 CARB state totals for anthropogenic emission sectors [CARB, 2014], with small (< 50 Gg CH_4/yr) adjustments for some regions and sectors (per ARB staff private communication). The spatial distribution of the dairy livestock emissions was revised by incorporating the 2012 county-level dairy statistics from USDA (http://www.nass.usda.gov/Statistics_by_State/California/Publications/County_Estimates/) to the spatial distribution from Jeong et al. [2013]. This revision changed the dairy livestock emissions for each region due to recent changes in the number of dairy cows, in particular for SoCAB. The current dairy livestock emissions in SoCAB (Table 2) decreased by ~50% compared to those (~80 Gg CH_4/yr) of Jeong et al. [2013], which was based on the 2004 statistics reported in Salas et al. [2009], reflecting the decrease in the number of dairy cows in the region (see Figure S1 in the Supporting Information (SI) for the trend of dairy cows in SoCAB). For natural wetlands, we used the prior emission map from Jeong et al. [2013].

Table 2 provides annual CALGEM prior emissions used in this study by source and region, and Figure 1 shows the annual total emission map for the CALGEM prior emission model along with the sub-region classification. The regions in this study are different from those in Jeong et al. [2013] and follow the California Air Basins (http://www.arb.ca.gov/ei/maps/statemap/abmap.htm). Inversion results are summarized by region to be compared with the prior emissions. Based on the prior emission estimates, the

Central Valley (Regions 3 and 8, Sacramento Valley (SV) and San Joaquin Valley (SJV), respectively) accounts for 55% of the total statewide CH_4 emissions and the two major urban regions (Regions 7 and 12, SFBA and SoCAB) account for 29% of the total. In terms of source sectors, livestock emissions represent 52% of the state total emission followed by landfills (20%) and natural gas (17%; petroleum production included). Livestock emissions are concentrated in Region 8 (San Joaquin Valley) where 86% (667 Gg CH_4 / 775 Gg CH_4) of the region's total emissions are from livestock. This is consistent with a recent study by Gentner et al. [2014] that suggests the majority of CH_4 emissions in the San Joaquin Valley are from dairy operations.

2.3. Atmospheric Transport Modeling

We used the coupled WRF-STILT (Weather Research and Forecasting and Stochastic Time-Inverted Lagrangian Transport) model for particle trajectory simulations [Lin et al., 2003; Skamarock et al., 2008; Nehrkorn et al., 2010]. The WRF-STILT model has been used to constrain GHG emissions in many studies including airborne measurement-based (e.g., Gerbig et al., [2003]; Kort et al., [2008]) and tower measurement-based (e.g., Zhao et al. [2009], Jeong et al. [2012a; 2012b; 2013], Newman et al. [2013]) applications. We adopt the set-up used in Jeong et al. [2013] to run the STILT model. In this set-up, an ensemble of 500 STILT particles are run backwards in time for 7 days driven with meteorology from the WRF model (version 3.5.1) [Skamarock et al., 2008]. Hourly predicted signals based on WRF-STILT are aggregated into 3-houly averages for inverse modeling.

The WRF model simulations closely follow those described in Jeong et al. [2012a; 2012b; 2013] with some modifications. Here, we use version 3.5.1 of the WRF model [Skamarock et al., 2008]. As in Jeong et al. [2013], we simulated meteorology for four different horizontal resolutions of 36, 12, 4, and two 1.3 km (vertical levels = 50) using initial and boundary meteorological conditions provided by the North American Regional Reanalysis (NARR) dataset [Mesinger et al., 2006]. In this study, the 1.3-km domain for the metropolitan area of Los Angeles was extended to better resolve outflow from the SoCAB region into eastern valleys that include the VTR site (see d04 in Figure 2). As in Jeong et al. [2013] we applied 2-way coupling between domains and 3-D analysis nudging at the outer domain every three hours using the NARR product.

For surface physics, we use two different land surface models (LSM) depending the location of each site as in Jeong et al. [2013]. For the Central Valley, we use the five-layer thermal diffusion LSM (5-L LSM) to account for irrigation in the land surface process during summer while using the Noah LSM [Chen and Dudhia, 2001] for other seasons. This is because the Noah LSM overestimates the planetary boundary layer (PBL) in the Central Valley without considering irrigation properly (dry surface leads to overestimation in PBL) [Jeong et al., 2013]. For the urban areas (e.g., SoCAB), we used the Noah LSM following Newman et al. [2013].

We also use different PBL schemes depending on the location of the GHG site. As a default for urban areas, we use the MYNN2 PBL scheme [Nakanishi and Niino, 2006] coupled with the Noah LSM. This is because we found that the MYJ scheme [Mellor and Yamada, 1982; Janjić, 1990] often underestimates nighttime PBL although it represents daytime PBL well. For the Central Valley region we also use the MYNN2 PBL scheme except for summer for which we used the MYJ scheme as in Jeong et al. [2013] coupled with the 5-L LSM for the Central Valley site. Some sites required improved representation of topographic influences on boundary layer meteorology during winter. Based on the transport evaluation using predicted and measured CO data, we apply the Yonsei University (YSU) scheme [Hong et al., 2006] with additional parameterization that corrects for surface wind biases at sites with complex topography (e.g., winter season in the southern San Joaquin Valley) [Jiménez and Dudhia, 2012].

A more complete evaluation of the WRF model simulations and transport errors associated are described in Bagley et al. (submitted to J. Geophy. Res.; henceforth, Bagley et al.), including a comparison of measured and predicted CO for the same period as this study (June 2013 – May 2014). The details for transport error are described in Bagley et al. using data from the wind profiler sites (Figure 2) and other observations. However, we note here that because CO is emitted from sources with different spatial and temporal variations than CH_4, the results of the CO comparison need to be interpreted with care in ascribing uncertainties to CH_4 emission estimates. In this study, we apply previous results from Jeong et al. [2013] to parameterize the prior probability distribution (instead of fixed values) for transport uncertainty and then update the prior uncertainty estimates using the hierarchical Bayesian method as described below.

2.4. Bayesian Inverse Model

We used a hierarchical Bayesian inversion (HBI, Ganesan et al. [2014]) method to estimate regional CH_4 emissions in California. In this work we develop an HBI method with more complex structure in representing the model-measurement mismatch matrix than Ganesan et al. [2014] for regional CH_4 emission quantification.

We start with Bayes' rule and describe each probability distribution in the hierarchical structure of parameters that include the scaling factor (a set of factors used to adjust prior emissions, denoted as $\boldsymbol{\lambda}$). Generally, Bayes' rule can be applied to multiple parameters at different levels as

$$p(\boldsymbol{\phi}, \boldsymbol{\theta}|\boldsymbol{D}) \propto p(\boldsymbol{D}|\boldsymbol{\phi}, \boldsymbol{\theta})p(\boldsymbol{\phi}, \boldsymbol{\theta}) \qquad (1)$$

$$= p(\boldsymbol{D}|\boldsymbol{\phi})p(\boldsymbol{\phi}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\theta}$ represent the generic parameters in vector form and **D** is data used to estimate the parameters. The first line in Equation 1 simply states the posterior probability is proportional to the likelihood function and prior distribution for the parameters. The re-factorization in the second line of Equation 1 holds because the data **D** depend only on the parameter $\boldsymbol{\phi}$ (thus $\boldsymbol{\theta}$ is factored out) and the values of $\boldsymbol{\phi}$ depend on the values of $\boldsymbol{\theta}$, constructing a hierarchical structure. The transition of $p(\boldsymbol{\phi}, \boldsymbol{\theta})$ to $p(\boldsymbol{\phi}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ is by the property of a conditional probability, given the dependence of $\boldsymbol{\phi}$ on $\boldsymbol{\theta}$. Any probabilistic model that can be factorized in chains as shown in Equation 1 is a hierarchical model [Kruschke, 2015].

The general model in Equation 1 can be applied to estimate surface emissions and their uncertainties. For GHG applications, the parameter vector $\boldsymbol{\phi}$ can be scaling factors for emission adjustment (or surface emission itself). The vector $\boldsymbol{\theta}$ can be a set of parameters including the hyper-parameters (e.g., mean) that determine the distribution for the scaling factor or surface emissions.

We use the following linear model for estimating scaling factors for regional emissions [Zhao et al., 2009; Jeong et al., 2012a; 2012b; 2013; Wecht et al., 2014]

$$\mathbf{y} = \mathbf{K}\boldsymbol{\lambda} + \mathbf{v} \qquad (2)$$

where **y** is the measurement vector ($n \times 1$), which represents 3-hourly local mixing ratio time series after subtracting background values, $\mathbf{K} = \mathbf{FE}$ (an $n \times k$ matrix), **F** is the footprint ($n \times m$), **E** is prior emissions ($m \times k$), $\boldsymbol{\lambda}$ is a $k \times 1$ vector for scaling factors with a covariance matrix **Q** ($k \times k$), and **v** is a vector representing the model-measurement mismatch with a covariance matrix **R** ($n \times n$). In this study we solve for a vector of 195 for $\boldsymbol{\lambda}$ which includes 0.3°×0.3° grid cells (a total of 183) within the major regions (i.e., Regions 3, 7, 8 and 12). We aggregated grid cells from other 12 regions at the sub-region scale so that the number of parameters can be reduced for those regions with low prior emissions and weak sensitivity to the measurement sites. Thus, after solving for $\boldsymbol{\lambda}$ using the HBI method and multiplying it by **E**, we can obtain posterior emissions (a vector of m).

For the model in Equation 2, the joint parameters we need to estimate are

$$\boldsymbol{\Theta} = \{\boldsymbol{\lambda}, \boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\sigma}_{\boldsymbol{\lambda}}, \boldsymbol{\sigma}_{\boldsymbol{R}}, \eta, \tau\} \qquad (3)$$

where $\boldsymbol{\lambda}$ is the scaling factor, $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ is the prior (i.e., hyper-parameter) mean for $\boldsymbol{\lambda}$, and $\boldsymbol{\sigma}_{\lambda}$ is the uncertainty for $\boldsymbol{\lambda}$ (i.e., square root of diagonal elements of **Q**). In HBI using a sampling method, $\boldsymbol{\lambda}$ is sampled from a probability distribution with mean $\boldsymbol{\mu}_{\boldsymbol{\lambda}}$ and standard deviation $\boldsymbol{\sigma}_{\lambda}$, which are also estimated (as part of $\boldsymbol{\Theta}$) instead of being prescribed as in previous work (e.g., Jeong et al. [2013], see below and SI for details). $\boldsymbol{\sigma}_{\mathbf{R}}$, η and τ are the parameters used to construct the model-measurement mismatch matrix **R** (see below for the representation of **R**). The diagonal elements of **R** represent the total model-measurement mismatch errors that are propagated through the inversion while **Q** is used to define the uncertainty level for the prior emission. These two quantities need to be either prescribed with known values or estimated. In HBI we estimate the joint parameter set simultaneously, using the measurements only once [Ganesan et al., 2014]. This joint estimation is different from previous approaches (e.g., Jeong et al. [2013]) where the covariance matrix **R** was prescribed via explicit estimation without using atmospheric measurements. It is also different from other methods where atmospheric measurements were used to optimize **R**, and measurements were thereafter also used for inversions (e.g., Michalak et al., [2005]).

With the parameter set identified, we need to write out the posterior probability up to the likelihood function and prior densities. We apply the identified joint parameter (i.e., **Θ**) to the general formulation of a hierarchical model in Equation 1 to express the posterior probability as

$$p(\boldsymbol{\lambda}, \boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\sigma}_{\boldsymbol{\lambda}}, \boldsymbol{\sigma}_{\boldsymbol{R}}, \eta, \tau | \boldsymbol{y}) \propto p(\boldsymbol{y} | \boldsymbol{\lambda}, \boldsymbol{\sigma}_{\boldsymbol{R}}, \eta, \tau) p(\boldsymbol{\lambda} | \boldsymbol{\mu}_{\boldsymbol{\lambda}}, \boldsymbol{\sigma}_{\boldsymbol{\lambda}}) p(\boldsymbol{\mu}_{\boldsymbol{\lambda}}) p(\boldsymbol{\sigma}_{\boldsymbol{\lambda}}) p(\boldsymbol{\sigma}_{\boldsymbol{R}}) p(\eta) p(\tau) \quad (4)$$

where the right-hand side shows the likelihood function and the prior distribution for each parameter. Note that in Equation 4 all variables are in vector form except for η and τ. To build Markov chain Monte Carlo (MCMC) samplers for the posterior distribution in Equation 4, the JAGS system (just another Gibbs sampler, Plummer [2003]) is used together with the R statistical language (https://cran.r-project.org/). JAGS has been widely used for statistical inference studies in many fields including ecology and genetics [Korner-Nievergelt et al., 2015; McKeigue et al., 2010]. The individual probability distributions (i.e., probability density functions) in Equation 4 are described below.

First, for the likelihood function we use

$$p(y|\boldsymbol{\lambda}, \boldsymbol{\sigma_R}, \eta, \tau) \sim N(\boldsymbol{K\lambda}, \boldsymbol{R}) \qquad (5)$$

where *N* is the normal distribution(here multivariate truncated normal, Miller et al. [2014]; Michalak [2008]) with mean $\mathbf{K}\lambda$ ($n \times 1$) and covariance $\mathbf{R}$ ($n \times n$). Note that $\mathbf{y}$ is conditionally independent of all other parameters given $\boldsymbol{\lambda}$, $\boldsymbol{\sigma}_\mathbf{R}$, η and τ.

In order to estimate parameter values with Bayesian inference, prior uncertainty needs to be specified. In the hierarchical model, we need to include prior uncertainty for the joint parameter set $\boldsymbol{\Theta}$ using a series of distributions. The scaling factor $\boldsymbol{\lambda}$ is sampled from a normal distribution instead of a fixed value (e.g., Jeong et al. [2013], Wecht et al., [2014]) as

$$p(\boldsymbol{\lambda}) \sim N(\boldsymbol{\mu_\lambda}, \boldsymbol{\sigma}_\lambda) \quad (6)$$

where $\boldsymbol{\mu_\lambda}$ itself is sampled from a truncated normal distribution [Miller et al., 2014; Michalak, 2008] with a mean of 1 and a standard deviation of 0.5 so that 68% of the samples are within 50 ~ 150% from the mean, which is a similar set-up to that of Ganesan et al. [2014]. $\boldsymbol{\sigma}_\lambda$ is modeled using a half Cauchy distribution, which is one of the recommended distributions for model variances [Gelman and Hill, 2007; Gelman et al., 2014; Korner-Nievergelt et al., 2015]. The hyper-parameterization ("hyper" meaning the upper level in the hierarchy) for $\boldsymbol{\sigma}_\lambda$ can formally be expressed as

$$\boldsymbol{\sigma_\lambda} \sim hCauchy\ (0,1) \quad (7)$$

where *hCauchy* is the half-Cauchy distribution. Note we take the absolute value from the Cauchy distribution so that we consider the positive values only (i.e., half Cauchy). Equation 7 suggests that if we generate random samples (large enough) from Equation 7 we get a median value close to 1. Thus, the use of 1 for the half Cauchy scale parameter (the larger the scale parameter, the more spread out the distribution) is similar to assuming the uncertainty for $\boldsymbol{\lambda}$ is 100% in the classical Bayesian inversion (e.g., Zhao et al., [2013], Jeong et al., [2013]). The difference is that in this study $\boldsymbol{\sigma}_{\lambda}$ is sampled from a distribution with a heavy tail (see Figure S2 in SI for an example half Cauchy distribution) so that $\boldsymbol{\sigma}_{\lambda}$ can be optimized from a broad distribution (instead of being a fixed value such as 50% of the mean emission).

For the model-measurement mismatch covariance matrix **R**, we use an exponential covariance function [Rasmussen and Williams, 2006]

$$R_{i,j} = \eta^2 \exp\left(-\frac{1}{\tau}|t_i - t_j|\right) + \delta_{i,j}\sigma_{R_s}^2 \quad (8)$$

where η, τ, and σ_{R_s} are parameters that define the covariance function, *t* is the measurement time, and δ is the Kronecker delta function (value of 1 if i = j, otherwise zero). We use two terms in Equation 8 to ensure the positive definiteness of **R** [Stan Development Team, 2015]. The second term in Equation (8) (i.e., with the Kronecker delta function) is analogous to the noise variance in the regression equation [Rasmussen and Williams, 2006]. Note that here we use the L_1 norm (i.e., $|t_i - t_j|$) as in Ganesan et al. [2014]. The subscript *s* in σ_{R_s} indicates that σ_R is estimated for each site as was done in Jeong et al. [2013] for their multi-tower analysis. This set of multiple parameters for σ_R adds more complexity to the model (than estimating a single value for σ_R) but also reflects the fact that model-measurement errors are not uniform across California.

We model σ_{R_S} using the half Cauchy distribution as in σ_{λ} [Gelman and Hill, 2007; Gelman et al., 2014; Korner-Nievergelt et al., 2015]. The scale parameter (in the hyper-parameter sense) for the half Cauchy distribution for σ_{R_S} is calculated using the first order approximation method following Jeong et al. [2012a; 2012b; 2013] and used as

$$p(\sigma_{R_s}) \sim hCauchy(0, \sigma_{R_{p|s}}) \quad (9)$$

where $\sigma_{R_{p|s}}$ is the first-order estimate for σ_{R_s} and includes errors from several sources (e.g., transport and background errors) combined in quadrature (see Text S1 in SI for details on first-order estimation for σ_{R_s}). More details for $\sigma_{R_{p|s}}$ are also described in the following section.

For η, we use non-informative prior as

$$\eta \sim unif(0, L) \quad (10)$$

where η is allowed to vary from 0 to L with an equal probability of 1/L. In this study we use $\sigma_{R_{p|s}}$ as an upper limit for L because in our choice of the covariance function η is estimated to be smaller than σ_R and this ensures the positive definiteness of the **R** covariance matrix, which is strictly checked in the JAGS sampler [Plummer, 2003; Version 3.4].

Following Ganesan et al. [2014], we use the exponential distribution for τ as

$$\tau \sim \exp(\frac{1}{\tau_p}) \quad (11)$$

where τ_p is the hyper parameter for τ, which is assumed to be 7 days (typical synoptic time scale for transport, Ganesan et al. [2014]).

2.5. Uncertainty Matrix

The posterior distribution in Equation 4 is used to generate MCMC samples for the parameters (i.e., $\mathbf{\Theta}$) in Equation 3, which include the components of the error covariance matrices **R** and **Q**. In other words, we estimate the model-measurement mismatch covariance matrix (i.e., **R**) and prior emission uncertainty (**Q**) simultaneously with $\boldsymbol{\lambda}$ and other parameters (using the measurements just once) instead of using fixed values. In the case of the **R** matrix, it can be estimated without assuming prior knowledge (e.g., uniform distribution) or from a simple assumption for the hyper-parameter as in Ganesan et al. [2014] where for the hyper-parameter of the variance component of **R** they used the sum of the fixed instrument uncertainty and the uncertainty associated with propagating the calibration scale (0.05 pmol/mol, respectively). Here we take a more informed approach by using site-specific values in constructing **R**, drawing on the model-measurement mismatch uncertainties reported in Jeong et al. [2013] for the sites included in that study: ARV, MAD, TRA, WGC and STB. For other sites, we estimated the model-measurement uncertainty for summer of 2013 following the method from Jeong et al. [2012a; 2012b; 2013] (see Text S1 in SI for details). For other seasons, we scaled the summer uncertainty estimates in proportion to the monthly background-subtracted mean mixing ratio signal. We use these uncertainty values (i.e., $\sigma_{R_{p|s}}$) as the hyper-parameter for σ_{R_S} in the covariance function for **R**. As described above, $\sigma_{R_{p|s}}$ is used as the scale parameter in the half Cauchy distribution in Equation 9 (see Table S1 for $\sigma_{R_{p|s}}$). As shown in Equation (8), the diagonal elements of **R** were then calculated as the sum of squares of σ_{R_S} and η.

Figure 3 shows the (optimized) posterior model-measurement mismatch uncertainty (i.e., diagonal elements of **R**) given the atmospheric measurements for several measurements sites that constrain the major emission regions (SV, SJV and SoCAB regions) (see Figure S3 in SI for correlation between posterior (σ_{R_S}) and prior ($\sigma_{R_{p|s}}$)). As described, the HBI approach allows for simultaneous estimation of model-measurement mismatch uncertainty values while inferring posterior emissions, using the measurements only once. This means that the model-measurement mismatch uncertainty has posterior estimates given the prior ($\sigma_{R_{p|s}}$) and data (i.e., model predictions and measurements). Overall, the posterior values follow the trend of the prior in seasonality and magnitude (Figure 3). In most sites, both the prior and posterior uncertainties are large during the winter season when boundary layer heights are low and predicted mixing ratios are very sensitive to the simulated boundary layer [Jeong et al., 2012a; 2013].

In our inverse model, the uncertainty in the prior emissions is expressed in terms of uncertainty in the scaling factors (i.e., $\boldsymbol{\sigma}_{\lambda}$, diagonal terms in $\mathbf{Q}$). Here, as with $\boldsymbol{\sigma}_{\mathbf{R}}$, the posterior values of $\boldsymbol{\sigma}_{\lambda}$ are also sampled from a half Cauchy distribution with a scale parameter of 1 (Equation 7). For the major emitting regions (3, 7, 8 and 12), the region average of prior uncertainties for individual pixels is estimated to be ~150% (see Figure S4 in SI), which is higher than the prescribed 70% in Jeong et al. [2013]. It is reasonable to expect this result because the pixel-based inversions have many more degrees of freedom and hence larger per pixel uncertainties than aggregate regions as in Jeong et al. [2013].

3. Results

3.1. State Total Emissions

State total emissions were estimated by optimizing 195 scaling factors each month (i.e., dimension of $\boldsymbol{\lambda} = 195 \times 1$) given the multi-site measurements and multiplying them by the CALGEM prior emissions, which were essentially the same as the CARB inventory at the sub-region scale (see Figure 1 for each sub-region). As described, we estimate a scaling factor for each 0.3° pixel within the major emission regions (i.e., SV, SFBA, SJV and SoCAB), which account for 84% of the CALGEM total emission. For other regions, we estimated a scaling factor for each region. Figure 4 compares predicted and background-subtracted measured mixing ratios using all data (used in the inversion) available for each season and also shows linear regression analysis results. Before inversion, the regression analysis estimates best-fit slopes to be 0.41 – 0.75 (predicted vs. measured). This simple analysis without full consideration of errors suggests that CH_4 emissions are underestimated by the CARB inventory. After inversion, the best-fit slope, root-mean-square error (RMSE) and coefficient of determination (r^2) for each season are significantly improved.

The HBI analysis estimates the state total annual emission is 2.04 - 2.90 Tg CH_4/yr at 95% confidence (median = 2.42) not including the (median) posterior estimate for natural wetlands (0.07 Tg CH_4/yr). This estimate is equivalent to 1.2 - 1.8 times the anthropogenic CH_4 emissions in CARB's current official inventory for the year 2013 (1.64 Tg CH_4/yr) [CARB, 2015]. Note that the state total in CARB's current official inventory for 2013 is only slightly different from the prior total in Table 2 after excluding the wetland emission. The state total emission estimate from HBI is consistent with the annual emission estimate from Jeong et al. [2013], 2.38±0.67 Tg CH_4/yr (at 95% confidence), which combined inverse model estimates for the Central Valley with urban emissions estimated by Wennberg et al. [2012].

As noted in the method section (Section 2), transport model error could affect the estimate of CH_4 emissions. Comparison of predicted and measured CO mixing ratios at the four towers during June 2013 – May 2014 (same period as this study) yields near-unity slopes for the majority of sites and seasons [Bagley et al.], suggesting that the WRF-STILT simulations are sufficient to estimate emissions of CO and likely other GHGs across California to within 10% ± 10% (at 95% confidence) on annual timescales. Based on this result, we add a mean transport bias uncertainty of 10% in quadrature to our Bayesian statistical uncertainty estimates to estimate total uncertainty in annual state total CH_4 emissions. After adding the transport bias uncertainty, we estimate state annual anthropogenic CH_4 emissions to be 2.42±0.49 Tg CH_4/yr (95%

confidence including transport bias uncertainty), higher than the anthropogenic emission in CARB's current inventory (1.64 Tg CH_4/yr in 2013). We note that the estimated CH_4 emissions drop to 1.0 - 1.6 times the CARB inventory if we correct for the 10% median CH_4 emissions assuming the bias in CO is applicable to CH_4. Undiagnosed sources of uncertainty may increase these error bounds beyond that indicated here. We also note that the transport error analysis based on CO rests on an assumption that *a priori* annual state total CO emissions are known to better than 10%, though Brioude et al. [2013] found that a comparison of measured and predicted CO (using the WRF-FLEXPART model) agreed to within about 15% for aircraft flights over SoCAB conducted in May and June 2010.

We estimate statewide CH_4 emissions for each season because our measurements are available for a full annual analysis (June 2013 – May 2014). This is the first analysis to estimate full seasonal CH_4 emissions using multi-tower measurements across California. Although Jeong et al. [2013] estimated seasonal CH_4 emissions in California using multi-tower measurements, they analyzed ten-month data only (not including July and August data) and did not constrain emissions from the southern California region. Figure 5 shows the estimated mean seasonal emissions for the state, which are the average of the monthly emissions belonging to the season. Note that the prior emissions in Figure 5 only partially account for seasonality because CALGEM has monthly emissions for crop agriculture (largely rice) and wetlands but not other sources. Across seasons, the posterior emissions are greater than the prior emissions without strong evidence for seasonality, similar to previous work by Jeong et al. [2013].

3.2. Emissions in Rural and Urban Regions

The hierarchical Bayesian inversion using multiple sites across California constrains CH_4 emissions from a significant portion of both rural and urban regions in California. In particular, the inverse analysis in this study yields a large reduction in the posterior uncertainty for the urban regions of California (e.g., SoCAB) compared to the inverse analysis by Jeong et al. [2013] where urban regions were under-sampled. We first examine the emissions for the rural regions of California, focusing on the Central Valley because it accounts for ~90% of the total rural emissions based on the CALGEM prior emission.

Figure 6 shows the comparison between prior and posterior emissions for the major emission regions that account for 84% of the state total in the CALGEM prior emission, including the Central Valley of California (see Table 3 for all regions). We estimate that the Central Valley (Regions 3 and 8) emissions are 1.02 – 1.74 Tg CH_4/yr (at 95% confidence, median = 1.38 Tg CH_4/yr). These estimates are consistent with the annual emission for the Central Valley estimated by Jeong et al. [2013], 1.57± 0.20 Tg CH_4/yr (95% confidence). Similarly, Wecht et al. [2014] estimated 1.23 Tg CH_4/yr for the Central Valley using a different transport model although it was only during the early summer period (May – June 2010). These results suggest emissions from the Central Valley are underestimated in the CALGEM prior emissions (0.94 Tg CH_4/yr). The spatial distribution of posterior emissions is shown in Figure 7 along with comparison with the CALGEM prior field. As can be seen in the figure, the posterior emissions for some of the pixels in the Central Valley are significantly larger than the prior. However, it should be noted that the uncertainty range for those pixels is also significantly large. This result shows that although the emissions at the sub-regional scale are well constrained in the Central

Valley (aggregated error at 95% confidence is ~25% of the posterior total of the Valley), the emission uncertainties for many of the individual pixels are still high. Bergamaschi et al. [2005] and Jeong et al. [2012a; 2013] reported that posterior emissions show anti-correlations between regions, suggesting that there could be some trade-offs of posterior emissions between regions. In this study, using pixel-based inversion for major emitting regions we have significantly reduced the anti-correlation in the posterior emissions at the sub-regional scale (e.g., between Region 3 and Region 7) to 0 - 20%, compared to those (up to 60% depending on the season) of Jeong et al. [2012a] (see Figure S5). This indicates that our total emission for each sub-region is much more independent than those of Jeong et al. [2012a; 2013].

For urban emissions of California, we focus on emissions from the two major urban regions (SoCAB and SFBA). According to the CALGEM prior, the two urban regions account for 25% of the state total emissions. The HBI analysis estimates the posterior emissions are 301 – 490 (median = 380, 95% confidence) Gg CH_4/yr for Region 12 (SoCAB), which are 0.9 – 1.4 times the prior (349 Gg/yr). This suggests that the prior inventory for SoCAB is consistent with our posterior estimate. Our posterior estimate is also consistent with the results of most of the recent studies that were conducted in SoCAB [Wunch et al., 2009; Hsu et al., 2010; Wennberg et al., 2012; Peischl et al., 2013; Wecht et al., 2014; Wong et al., 2015; Cui et al., 2015]. Figure 8 shows the comparison of estimated CH_4 emissions for SoCAB among eight different recent studies including this study. The estimate (600 Gg CH_4/yr) by Wunch et al. [2009] using the CH_4/CO_2 ratio is likely the upper limit for SoCAB CH_4 emissions and is not included in this comparison. Although the estimated emissions are consistent among the different studies given the reported uncertainty, there are some differences in the mean/median estimates. These differences may arise from different assumptions and undiagnosed uncertainties (e.g., spatial distribution of bottom-up emissions, transport model errors, different seasonal coverage). For example, most of the studies in SoCAB rely on the combination of measured CH_4 to CO_2 or CO ratios and the bottom-up inventory of CO_2 or CO, with uncertainties that assume those inventories are relatively well-known (e.g., 10% uncertainty assumption in CO_2 inventory by Wong et al. [2015]).

This study constrains CH_4 emissions for SFBA with a significant reduction in the posterior uncertainty, compared to Jeong et al. [2012; 2013]. We estimate the posterior emissions for SFBA to be 159 – 340 (median = 245) Gg CH_4/yr (at 95% confidence). These emission estimates are consistent with those reported by Fairley and Fischer [2015] where they reported a total of 240±60 Gg/yr (at 95% confidence) for the recent period of 2009 to 2012 using CH_4:CO enhancement ratios from 14 air quality sites in SFBA. For SFBA, we have two bottom-up estimates to be compared with our inverse analysis: CALGEM emission model (143 Gg CH_4/yr, see Table 2) and the Bay Area Air Quality Management District (BAAQMD) inventory (126 Gg CH_4/yr, [BAAQMD, 2015]). Compared to bottom-up estimates, actual CH_4 emissions in the SFBA are likely 1.1 – 2.4 and 1.3 – 2.7 times larger than the CALGEM prior and BAAQMD's inventory, respectively, suggesting that both inventories are lower than our posterior estimate.

3.3. Source Attribution of Emissions

We investigate the likely sources of emissions in the rural and urban regions of California. We estimate CH_4 emissions from different sources assuming the spatial distribution of the CALGEM emission model. Based on this assumption, we scale individual sector prior emissions at each pixel or region by the inferred scaling factors from the HBI analysis. Figure 9 (left) shows posterior annual emissions for the HBI analysis by sector. The posterior emissions (804 – 1410 Gg CH_4/yr, median = 1070 Gg) for the dairy livestock (DLS) are 1.1 - 1.9 times larger than the prior emissions. Assuming the distribution of the prior, the posterior estimates for the non-dairy livestock (199 – 345 Gg CH_4/yr, median = 263 Gg) are also 1.3 – 2.2 times larger than the prior. The combined total emissions for dairy and non-dairy livestock emissions (1050 - 1699 Gg CH_4/yr) are 1.2 – 1.9 times higher than the CALGEM prior. The underestimate in livestock emissions agrees with the results described in the region analysis that posterior emissions in the Central Valley (Regions 3 and 8) are larger than the CALGEM prior. This is also consistent with the reported livestock emissions (1265 – 1805 Gg CH_4/yr, at 95% confidence) by Jeong et al. [2013]. A recent global study suggests a similar underestimation for manure management in a bottom-up inventory. Based on published data on field-scale measurements of GHG emissions, Owen and Silver [2015] report that predicted CH_4 emissions by the Intergovernmental Panel on Climate Change (IPCC) Tier 2 method are lower than the mean estimates using the field measurements for most manure management practices. However, we caveat the source attribution above because the spatial distribution of sources by sector may not be perfectly captured in the CALGEM model. In terms of seasonality by sector, Figure 9 (right) suggests that except for WL and CP, the seasonal variation in the emissions is small, showing similar seasonal posterior emissions within error (Figure 9).

Our inverse analysis also suggests that actual natural gas (NG; includes petroleum production) and landfill (LF) emissions are likely larger than the prior emissions. Our posterior NG emissions (305 – 502 Gg CH_4/yr) are higher than the prior (283 Gg, see Figure 9) used in this study but consistent with that (331 Gg CH_4/yr) estimated by Jeong et al. [2014] where they find their spatially explicit bottom-up inventory for NG itself is generally lower than those of top-down analyses (e.g., Peischl et al. [2013], Wennberg et al. [2012]). The result for seasonal emissions by sector in Figure 9 (right) shows that the seasonal variation for NG and LF is small, consistent among seasons within error. Other sources, including petroleum refining and mobile (RM), wastewater (WW), crop (rice) emissions (CP), and wetlands (WL) are generally similar between prior and posterior emissions. The rice emissions are 39 – 101 Gg CH_4/yr (at 95% confidence), which are consistent with those of Jeong et al. [2013] (68±18 Gg, at 95% confidence) and Peischl et al. [2012] (~85 Gg).

4. Discussion and Conclusions

We further discuss likely source emissions by comparing our estimates with results from previous studies. Jeong et al. [2013] estimated annual CH_4 emissions from the livestock source sector in the San Joaquin Valley (Region 8) to be 1.13 ± 0.42 Tg CH_4/yr (at 95% confidence), significantly higher than all other sources combined in the region. This is consistent with the finding by Gentner et al. [2014] who concluded that the “vast majority” of the total emission in San Joaquin Valley is due to dairy operations. In another similar study, Guha et al. [2015] used collocated measurements of CO and various volatile organic compounds (VOCs, e.g. alkanes) and a Positive Matrix Factorization (PMF) technique to estimate the contribution of regional sources to observed enhancements of CH_4. The results in Guha et al. [2015] indicate that the livestock emissions account for a majority of the CH_4 (70 - 90%, uncertainty = 29%) enhancements based on measurements near Bakersfield, California during May - June 2010. The reported 29% uncertainty is calculated from the standard deviation in the mass fraction of CH_4 attributed to the dairy source factor profile as estimated from a bootstrapping method. Although these two studies do not report estimated emissions by mass, they suggest a significant portion of the total CH_4 emission in the San Joaquin Valley (Region 8) is attributed to the livestock sector.

More quantitatively, Jeong et al. [2014] estimated CH_4 emissions from the natural gas sector (petroleum production included) for the state based on activity data and reported emission factors (mostly from US Environmental Protection Agency (EPA)). They estimated the emission from the natural gas sector to be 128 Gg CH_4/yr for the San Joaquin Valley, the majority of which was from petroleum and natural gas production. After adjusting this bottom-up estimate based on the result in SoCAB by Peischl et al. [2013], they estimated the natural gas emission in San Joaquin Valley to be 162.6 Gg CH_4/yr, with the San Joaquin Valley accounting for 30% of the state total natural gas emissions. The adjusted natural gas emission (i.e., 162.6 Gg) by Jeong et al. [2014] is 11 - 19% of the annual total emissions (0.86 – 1.49 Tg CH_4) in the San Joaquin Valley estimated in this study, which is consistent with Gentner et al. [2014], Guha et al. [2015] and Jeong et al. [2013]. Note that, based on the CALGEM prior, the San Joaquin Valley emits 82% of the total CH_4 emissions in the Central Valley, 86% of which is from the livestock sector. These results suggest that our a priori assumption about the ratio of livestock emissions to the total in the San Joaquin Valley is likely similar to the source attribution of the actual emissions in Region 8. Furthermore, our source analysis indicates that the posterior emissions for landfill, natural gas, and wastewater are generally consistent with or slightly higher than our CALGEM prior, and livestock emissions are higher than the prior although this is a statewide result (see Figure 9(left)). Given this source analysis result, the higher posterior emissions in San Joaquin Valley (Region 8) from our region analysis (1.1 – 1.9 times the CALGEM prior) are likely mainly due to livestock sources.

We also examine the emissions in SoCAB for possible source attributions by combining the results from this study and other previous work. In this study we estimated that the CH_4 emissions in SoCAB are 330 – 421 Gg CH_4/yr (median = 380, here we report the 68% confidence interval for comparison with other work). Combining the recent studies in SoCAB including this study (for Wunch et al. the estimate based on CO/CH_4 ratios is used, see Figure 8) we estimate the SoCAB CH_4 emission is 341 - 465 Gg CH_4/yr (at 95% confidence, mean/median = 403 Gg CH_4) [Wunch et al., 2009; Hsu et al., 2010; Wennberg et al. 2012; Peischl et al., 2013; Wong et al., 2015; Wecht et al., 2014; Cui et al., 2015]. To calculate the uncertainty in this estimate, we generated 50000 MCMC samples for each study based on the mean and uncertainty reported in individual studies (similar to generating samples for the prior distributions in HBI) and combined them for an overall mean distribution. Note that the uncertainty for the overall mean is smaller than those of the individual studies because the mean estimates of individual studies are close to the combined mean (i.e., a small spread around 403 Gg CH_4), suggesting emission estimates in SoCAB are converging among different studies. It should also be noted that the emission estimates for SoCAB in most of these previous studies including ours include emissions from petroleum seepage and abandoned wells in the total without distinguishing these as non-anthropogenic emissions. This suggests that the CALGEM prior total for SoCAB (349 Gg CH_4) scaled by the CARB inventory is comparable to the recent top-down estimates for SoCAB. For source attribution, Wennberg et al. [2012] suggest that the majority of the CH_4 enhancements observed are likely due to natural gas activities, while Peischl et al. [2013] estimates 192± 54 Gg CH_4 for the combination of emissions from natural gas transmission and distribution plus local seeps, and 32±7 Gg CH_4 for oil and gas production and processing. Hence the total of fossil fuel related activities from Peischl et al. [2013] is 224 ±55 Gg CH_4, assuming uncorrelated errors in the above estimates. This estimate is larger than our CALGEM prior for the combined total from the natural gas (NG) and refining and on-road mobile (RM) sectors of 124 Gg (see Table 2) by a factor of 1.4 – 2.3, suggesting an underestimate for total fossil fuel related emissions in the CALGEM prior for SoCAB. Lyon et al. [2015] reported a similar result in a recent sub-regional scale study for the Barnett Shale region where they estimated higher CH_4 emissions from the oil and gas sector than three inventories by factors of 1.5 – 4.3. For landfill, wastewater and livestock sectors, the CALGEM prior estimates 224 Gg CH_4/yr for SoCAB, which is consistent with that (182 ± 54 Gg CH_4/yr) of Peischl et al. [2013]. For livestock, Cui et al. [2015] estimates emissions in SoCAB to be 52±15 Gg CH_4/yr, which is consistent with the CALGEM prior (44 Gg CH_4/yr). Last, Cui et al. [2015] also estimated a combined CH_4 emission of 347±71 Gg CH_4/yr for the landfill and natural gas sectors. This also indicates that natural gas emissions are likely larger than the CALGEM natural gas prior, because their minimum estimate (276 Gg) for the landfill and natural gas sectors is larger than that of the CALGEM prior for natural gas and landfills together (268 Gg). Taken together, these results suggest that while the prior emissions (SoCAB total of 349 Gg) are towards the low end of the top-down estimates (341 – 465 Gg), underestimation in NG emissions from the CALGEM prior model is possible as indicated by the higher top-down estimates from Peischl [2013] and Cui [2015].

In summary, our measurement network across California constrains CH_4 emissions from California's urban and rural emissions, and the added measurement sites to the CH_4 network significantly reduced the posterior uncertainty estimates. This suggests that the inverse framework based on the measurement network can be an effective approach to quantifying emissions at the regional scale and monitoring long-term spatial and temporal changes in emissions. Although the CO comparison [Bagley et al.] appears largely consistent with expectation, it is possible that undiagnosed sources of error affect the CH_4 emission estimates. In the future, a combination of improved prior emission and meteorological models, expanded multi-gas measurements, and inverse model analyses will reduce uncertainty in California's GHG emissions. Also, more efforts are needed to constrain emissions by both sector and region. For example, while our results and other studies indicate both livestock and natural gas emissions appear to be underestimated, attribution of the magnitude of errors to specific sectors is difficult. A recent study on CH_4 emissions from the Aliso Canyon blowout in Los Angeles emphasizes the utility of tracers (e.g., ethane) for source speciation [Conley et al., 2016]. Using both methane and ethane measurements, Conley et al. [2016] reported that at its peak the Aliso Canyon event doubled SoCAB emissions during the 3-month period, producing a total of 97 Gg CH_4, which is 28% of the SoCAB total CH_4 emission (349 Gg/yr) from our CALGEM prior model. Given the importance of distinguishing the regional variations in dominant CH_4 sources (e.g., Central Valley vs. SoCAB) and large-scale events such as the Aliso Canyon blowout, a combination of facility specific emission measurements and regionally representative measurements of source-specific tracers (e.g., CO, VOCs, and potentially CH_4 isotopes) [Townsend-Small et al., 2012; Peischl et al., 2013; Guha et al., 2015] are likely to prove useful in the future.

Acknowledgements

We thank Dave Field, Dave Bush, Edward Wahl, Ken Reichl, Toby Walpert, and particularly Jon Kofler for assistance with measurements at WGC and analysis of data from radar wind profiler sites, Christina Harth for assistance with measurements at THD, John Lin, Christoph

Gerbig, Steve Wofsy, Janusz Eluszkiewicz, Thomas Nehrkorn for sharing the STILT code and advice, Anita Ganesan for the motivation of the HBI approach, Chris Potter and William Salas for sharing modeled CH_4 emission for use as a priori estimates, Ed Dlugokencky and Colm Sweeney for sharing data for CH_4 background estimates, Bart Croes, Jorn Herner, Abhilash Vijayan, Matthias Falk, Richard Bode, Anny Huang, Jessica Charrier, Kevin Eslinger, Larry Hunstaker, Ken Stroud, Mac McDougall, Jim Nyarady, and others for sharing CARB emissions information and providing valuable review comments, and Krishna Muriki for assistance running the WRF-STILT models on the LBNL-Lawrencium cluster. The data used in the inversion are in supplements (Figure S9 and S10) and the CALGEM prior emission distribution is available at http://calgem.lbl.gov/. This study was supported by the University of California's Discovery Grant Program and the California Air Resources Board Research Division under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

References

Andrews, A. E., et al. (2014), CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, Atmos. Meas. Tech., 7, 647–687, doi:10.5194/amt-7-647-2014 .

BAAQMD (2015), Bay Area Emissions Inventory Summary Report: Greenhouse Gases Base Year 2011 (updated January 2015), http://www.baaqmd.gov/~/media/Files/Planning%20and%20Research/Emission%20Inventory/BY2011_GHGSummary.ashx?la=en (accessed May 2015)

Bagley, J., et al., Assessment of an atmospheric transport model for annual inverse estimates of California greenhouse gas emissions (submitted to J. Geophys. Res. Atmos.; Manuscript # 2016JD025361).

Bergamaschi, P., M. Krol, F. Dentener, A. Vermeulen, F. Meinhardt, R. Graul, M. Ramonet, W. Peters, and E. J. Dlugokencky (2005), Inverse modelling of national and European CH_4 emissions using the atmospheric zoom model TM5, Atmos. Chem. Phys., 5, 2431–2460, doi:10.5194/acp- 5-2431-2005.

Brioude, J., et al. (2013), Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NOx and CO_2, *Atmospheric Chemistry and Physics*, 13, 3661-3677. doi:10.5194/acp-13-3661-2013.

CARB (2014), California Greenhouse Gas Emissions Inventory. California Air Resources Board Staff Report, Accessed January 2015 (http://www.arb.ca.gov/cc/inventory/inventory.htm, version March 2014).

CARB (2015), California Greenhouse Gas Emissions Inventory. California Air Resources Board Staff Report, Accessed September 2015 (http://www.arb.ca.gov/cc/inventory/inventory.htm, version April 2015).

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface hydrology model with the Penn State NCAR MM5 modeling system. Part 1: model implementation and sensitivity, Mon. Weather Rev., 129, 569–585.

Conley, S., G. Franco, I. Faloona, D. R. Blake, J. Peischl, and T. B. Ryerson (2016), Methane emissions from the 2015 Aliso Canyon blowout in Los Angeles, CA, Science, 10.1126/science.aaf2348.

Cui, Y. Y., J. Brioude, S. A. McKeen, W. M. Angevine, S.-W. Kim, G. J. Frost, R. Ahmadov, J. Peischl, N. Bousserez, Z. Liu, T. B. Ryerson, S. C. Wofsy, G. W. Santoni, E. A. Kort, M. L. Fischer, and M. Trainer (2015), Top-down estimate of methane emissions in California using a mesoscale inverse modeling technique: The South Coast Air Basin. J. Geophys. Res. Atmos., 120, 6698–6711. doi: 10.1002/2014JD023002.

Fairley, D., and M. L. Fischer (2015). Top-down methane emissions estimates for the San Francisco Bay Area from 1990 to 2012, Atmospheric Environment, doi:10.1016/j.atmosenv.2015.01.065.

Ganesan, A. L. et al. (2014), Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods. Atmos. Chem. Phys., 14, 3855–3864, doi:10.5194/acp-14-3855-2014.

Gelman, A. and J. Hill (2007), Data analysis using regression and multilevel/hierarchical models, Cambridge University Press, New York, NY, USA.

Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2014), Bayesian data analysis (3rd Ed.), Chapman & Hall/CRC.

Gentner, D.R., et al. (2014), Emissions of organic carbon and methane from petroleum and dairy operations in California's San Joaquin Valley, Atmos. Chem. Phys. 13, 28225–28278.

Gerbig, C, J. Lin, S. Wofsy, B. Daube, A. E. Andrews, B. Stephens, P. S. Bakwin and C. Grainger (2003), Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework. J. Geophys. Res., 108(D24), doi:10.1029/2003JD003770.

Guha, A., D.R. Gentner, R.J. Weber, R. Provencal, A. Gardner, and A.H. Goldstein (2015), Source apportionment of methane and nitrous oxide in California's San Joaquin Valley at CalNex 2010 via positive matrix factorization, Atmos. Chem. Phys., 15, 12043–12063, doi:10.5194/acp-15-12043-2015.

Hall, B. D., et al. (2014), Results from the International Halocarbons in Air Comparison Experiment (IHALACE), Atmos. Meas. Tech., 7(2), 469 – 490.

Hong, S., Y. Noh, and J. Dudhia (2006), A new vertical diffusion pack- age with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, doi:10.1175/MWR3199.1.

Hsu, Y.-K., T. VanCuren, S. Park, C. Jakober, J. Herner, M. FitzGibbon, D. R. Blake, and D. D. Parrish (2010), Methane emissions inventory verification in southern California, Atmos. Environ., 44, 1 – 7, doi:10.1016/j.atmosenv.2009.10.002.

Janjić, Z. I. (1990), The step-mountain coordinate: Physical package, Mon. Weather Rev., 118, 1429–1443.

Jeong, S., C. Zhao, A. E. Andrews, L. Bianco, J. M. Wilczak, and M. L. Fischer (2012a), Seasonal variation of CH_4 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 N_2O emissions from central California. Geophys. Res. Lett., 39, L16805, doi:10.1029/2012GL052307.

Jeong, S., Y.-K. Hsu, A. E. Andrews, L. Bianco, P. Vaca, J. M. Wilczak, and M. L. Fischer (2013), A multitower measurement network estimate of California's methane emissions, J. Geophys. Res. Atmos., 118, 11,339–11,351, doi:10.1002/jgrd.50854.

Jeong, S., D. Millstein, and M. L. Fischer (2014), Spatially explicit methane emissions from petroleum production and the natural gas system in California, Environ. Sci. Technol., 48, 5982–5990.

Jiménez, P A., and J. Dudhia (2012), Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model., J. Appl. Meteor. Climatol., 51, 300–316, doi: http://dx.doi.org/10.1175/JAMC-D-11-084.1.

Korner-Nievergelt, F, T. Roth, S. von Felten, J. Guélat, B. Almasi, and P. Korner-Nievergelt (2015), Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan: Including Comparisons to Frequentist Statistics. Elsevier Science.

Kort, E. A., J. Eluszkiewicz, B. B. Stephens, J. B. Miller, C. Gerbig, T. Nehrkorn, B. C. Daube, J. O. Kaplan, S. Houweling, and S. C. Wofsy (2008), Emissions of CH_4 and N_2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, Geophy. Res. Lett., 35, L18808, doi:10.1029/2008GL034031.

Kruschke, J. K. (2015), Doing Bayesian data analysis (2nd Ed.), Academic Press, 759 pp.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and C. A. Grainger (2003), A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model. J. Geophys. Res., 108(D16), 4493, doi:10.1029/2002JD003161.

Lyon, D. R., et al. (2015), Constructing a spatially resolved methane emission inventory for the Barnett Shale Region, Environ. Sci. Technol., 49, 8147−8157, DOI: 10.1021/es506359c.

McKeigue, P. M., H. Campbell, S. Wild, V. Vitart, C. Hayward, I. Rudan, A. F. Wright, and J. F. Wilson (2010). Bayesian methods for instrumental variable analysis with genetic instruments ('Mendelian randomization'): example with urate transporter SLC2A9 as an instrumental variable for effect of urate levels on metabolic syndrome, Int. J. Epidemiol, 39(3), 907-18. doi: 10.1093/ije/dyp397.

Mellor, G. L., and T. Yamada (1982), Development of a turbulence closure model for geophysical fluid problems, Rev. Geophys. Space Phys., 20, 851–875.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi (2006), North American Regional Reanalysis. Bull. Amer. Meteor. Soc., 87 (3), 343-360.

Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, J. Geophys. Res., 110, D24107, doi:10.1029/2005JD005970.

Michalak, A. M. (2008), A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, Water Resour. Res., 44, W09437, doi:10.1029/2007WR006645.

Miller, S. M., A. M. Michalak, and P. J. Levi (2014), Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geosci. Model Dev., 7, 303–315, doi:10.5194/gmd-7-303-2014.

Nakanishi, M. and H. Niino (2006), An improved Mellor Yamada level-3 model: its numerical stability and application to a regional prediction of advection fog, Bound. Layer Meteorol., 119, 397– 407.

Nehrkorn, T., J. Eluszkiewicz, S. C. Wofsy, J. C. Lin, C. Gerbig, M. Longo, and S. Freitas (2010), Coupled weather research and forecasting - stochastic time-inverted lagrangian transport (WRF-STILT) model. Meteor. Atmos. Phys., 107 (1), 51-64, doi:10.1007/s00703-010-0068-x.

Newman, S., S. Jeong, M. L. Fischer, X. Xu, C. L. Haman, B. Lefer, S. Alvarez, B. Rappenglueck, E. A. Kort, A. E. Andrews, J. Peischl, K. R. Gurney, C. E. Miller, and Y. L. Yung (2013), Diurnal tracking of anthropogenic CO_2 emissions in the Los Angeles basin megacity during spring 2010, Atmos. Chem. Phys., 13, 4359–4372, doi:10.5194/acp-13-4359-2013.

Owen, J., and W. L. Silver (2015), Greenhouse gas emissions from dairy manure management: a review of field-based studies, Global Change Biology, 21, 550–565, doi: 10.1111/gcb.12687.

Peischl, J., et al. (2012), Airborne observations of methane emissions from rice cultivation in the Sacramento Valley of California, J. Geophys. Res., 117, D00V25, doi:10.1029/2012JD017994.

Peischl, J., et al. (2013), Quantifying sources of methane using light alkanes in the Los Angeles basin, California, J. Geophys. Res. Atmos., 118, doi:10.1002/jgrd.50413.

Plummer, M. (2003), JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X.

Prinn, R. G., et al. (2000), A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, J. Geophys. Res. Atmos., 105(D14), 17751–17792.

Rasmussen, C. E., and C. K. I. Williams (2006), Gaussian Processes for Machine Learning, MIT Press, Cambridge, Massachusetts.

Salas, W., C. Li, F. Mitloehner, and J. Pisano (2009), Developing and applying process-based models for estimating greenhouse gas and air emissions from California dairies, Rep. CEC-500-2008-093, Public Interest Energy Res. Program, Calif. Energy Comm., Sacramento, Calif.

Skamarock, W.C., J. B. Klemp, J. Dudhia, D.O. Gill, D. M. Barker, X. Z. Huang, W. Wang, and J. G. Powers (2008), A description of the advanced research WRF version 3. Technical Note 475+STR. Mesoscale and Microscale Meteorology Division, NCAR, Boulder, Colorado.

Stan Development Team (2015), Stan Modeling Language: User's Guide and Reference Manual (Version 2.9.0.; http://mc-stan.org/documentation/).

Townsend-Small, A., S. C. Tyler, D. E. Pataki, X. Xu, and L. E. Christensen (2012), Isotopic measurements of atmospheric methane in Los Angeles, California, USA: Influence of "fugitive" fossil fuel emissions, J. Geophys. Res., 117, D07308, doi:10.1029/2011JD016826.

Wecht, K. J., Jacob, D. J., Sulprizio, M. P., Santoni, G. W., Wofsy, S. C., Parker, R., Bösch, H., and Worden, J. (2014) Spatially resolving methane emissions in California: constraints from the CalNex aircraft campaign and from present (GOSAT, TES) and future

(TROPOMI, geostationary) satellite observations, Atmos. Chem. Phys., 14, 8173-8184, doi:10.5194/acp-14-8173-2014.

Wennberg, P. O., W. Mui, D Wunch, E. A. Kort, D. R. Blake, E. L. Atlas, G. W. Santoni, S. C. Wofsy, G. S. Diskin, S. Jeong, and M. L. Fischer (2012), On the Sources of Methane to the Los Angeles Atmosphere. Environ. Sci. Technol., 46 (17), 9282 - 9289, doi:10.1021/es301138y.

Wong, K. W., Fu, D., Pongetti, T. J., Newman, S., Kort, E. A., Duren, R., Hsu, Y.-K., Miller, C. E., Yung, Y. L., and Sander, S. P. (2015), Mapping CH_4 : CO_2 ratios in Los Angeles with CLARS-FTS from Mount Wilson, California, Atmos. Chem. Phys., 15, 241-252, doi:10.5194/acp-15-241-2015.

Wunch, D., P. O. Wennberg, G. C. Toon, G. Keppel-Aleks, and Y. G. Yavin (2009), Emissions of greenhouse gases from a North American megacity, Geophys. Res. Lett., 36, L15810, doi:10.1029/2009GL039825.

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.

Tables

Table 1. GHG Sites Information across California

Site	Location	Latitude	Longitude	Inlet Height (m, a.g.l.)*	Data Availability
ARV	Arvin	35.24	-118.79	10	June 2013 – May 2014
CIT	Caltech, Pasadena	34.14	-118.12	10	June 2013 – May 2014
LVR	Livermore	37.67	-121.71	27	June 2013 – May 2014
MAD	Madera	36.87	-120.01	10	June 2013 – May 2014
STB	Sutter Buttes	39.21	-121.82	10	June 2013 – May 2014
STR	San Francisco	37.76	-122.45	232	June 2013 – May 2014
THD	Trinidad Head	41.05	-124.15	20	June 2013 – August 2013
TRA	Tranquility	36.63	-120.38	10	June 2013 – April 2014
TSB	Tuscan Buttes	40.26	-122.09	10	June 2013 – May 2014
VTR	Victorville	34.61	-117.29	90	June 2013 – August 2013
WGC	Walnut Grove	38.27	-121.49	91	June 2013 – May 2014
SBC	San Bernardino	34.09	-117.31	58	June 2013 – May 2014
SIO	Scripps Institution of Oceanography	32.87	-117.26	10	June 2013 – May 2014

*Inlet heights used in the inversion

Table 2. Annual CALGEM CH_4 Emissions by Region and Sector (Gg CH_4)

Source[a]\ Region[b]	GBV (6)	LC (5)	LT (15)	MC (4)	MD (10)	NC (2)	NCC (9)	NEP (1)	SoCAB (12)	SCC (11)	SD (14)	SFBA (7)	SJV (8)	SS (13)	SV (3)	Total
DLS	0.1	0.0	0.1	2.5	21.1	21.7	2.8	2.2	37.9	1.0	2.6	14.3	598.1	3.8	30.1	**738.3**
LF	1.1	1.5	0.0	2.6	8.2	2.3	9.0	1.1	157.0	14.3	26.3	53.9	28.7	3.1	26.4	**335.4**
NDLS	2.0	0.4	0.1	8.3	3.5	8.2	5.2	13.3	5.8	8.3	1.9	10.5	68.5	1.6	19.8	**157.4**
NG	0.2	0.4	0.2	2.3	5.5	1.7	4.0	0.4	112.2	17.4	16.2	38.8	51.1	3.1	29.8	**283.3**[c]
RM	0.1	0.1	0.0	2.1	2.5	0.4	0.9	0.1	12.0	1.2	2.2	10.0	4.3	0.5	3.3	**39.7**
WW	0.0	0.1	0.0	0.5	1.0	0.3	1.6	0.4	23.6	13.1	2.8	11.0	9.0	0.9	2.8	**67.1**
WL	0.3	0.0	0.0	0.5	0.1	0.2	0.2	9.9	0.9	0.5	0.2	4.1	14.0	0.1	7.1	**38.1**
CP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.5	0.0	47.8	**49.2**
Total	**3.8**	**2.4**	**0.5**	**18.8**	**41.8**	**34.8**	**23.7**	**27.5**	**349.3**	**55.8**	**52.2**	**142.5**	**775.2**	**13.2**	**167.0**	**1708.6**

[a]Sectors include dairy livestock (DLS), landfill (LF), non-dairy livestock (NDLS), natural gas including petroleum production and local processing (NG), petroleum refining and mobile sources (RM), wastewater (WW), wetland (WL), and crop (CP, largely rice).

[b]The number in the parentheses shows the region number shown in Figure 1.

[c]includes 24 Gg CH_4/yr from petroleum seeps (CARB staff private communication).

Table 3. Posterior Annual Emission Estimates (Gg CH_4/year) by Region

Regions[a]	1 (NEP)	2 (NC)	3 (SV)	4 (MC)	5 (LC)	6 (GBV)	7 (SFBA)	8 (SJV)	9 (NCC)	10 (MD)	11 (SCC)	12 (SoCAB)	13 (SS)	14 (SD)	15 (LT)
Prior	28	35	167	19	2	4	143	775	24	42	56	349	13	52	1
HBI Posterior (Upper)[b]	186	144	360	84	20	23	340	1486	180	243	162	490	68	145	2
HBI Posterior (Lower)[c]	1	1	164	1	0	0	159	859	1	1	1	301	1	37	0

[a]Region abbreviations are shown in the parentheses.
[b]97.5th percentile
[c]2.5th percentile

Figures

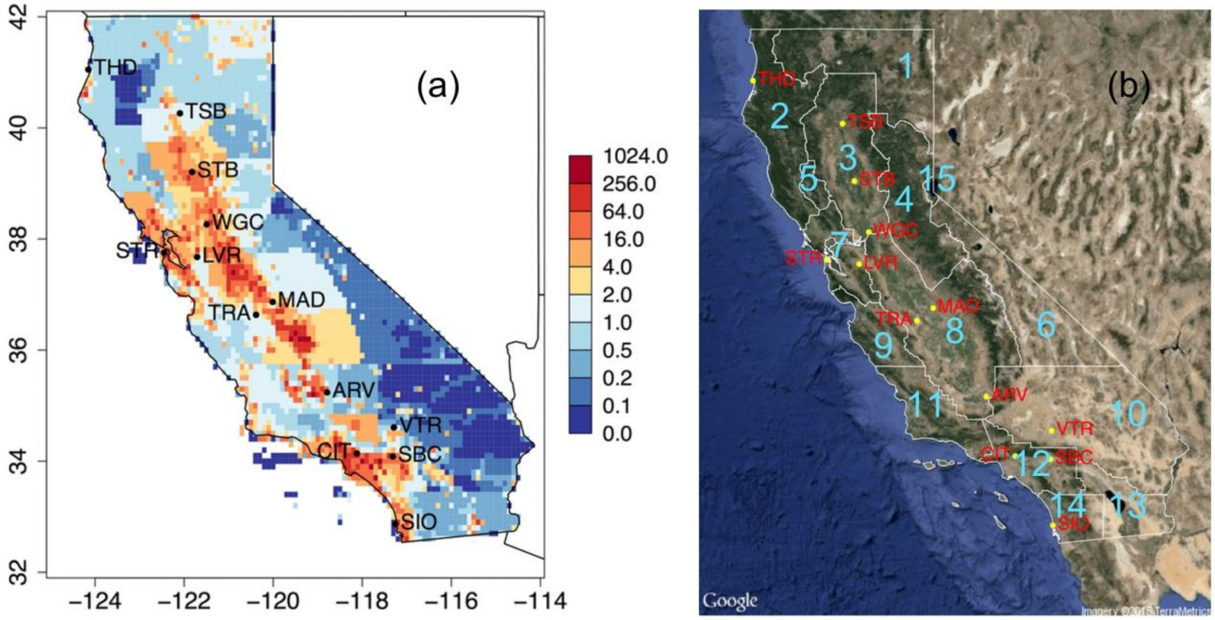

Figure 1. (a) CALGEM total (1.7 Tg CH_4/yr, 1 Tg = 10^{12} g) prior emissions (nmol/m^2/s) with locations of measurement sites across California, and (b) region classification (California air basins).

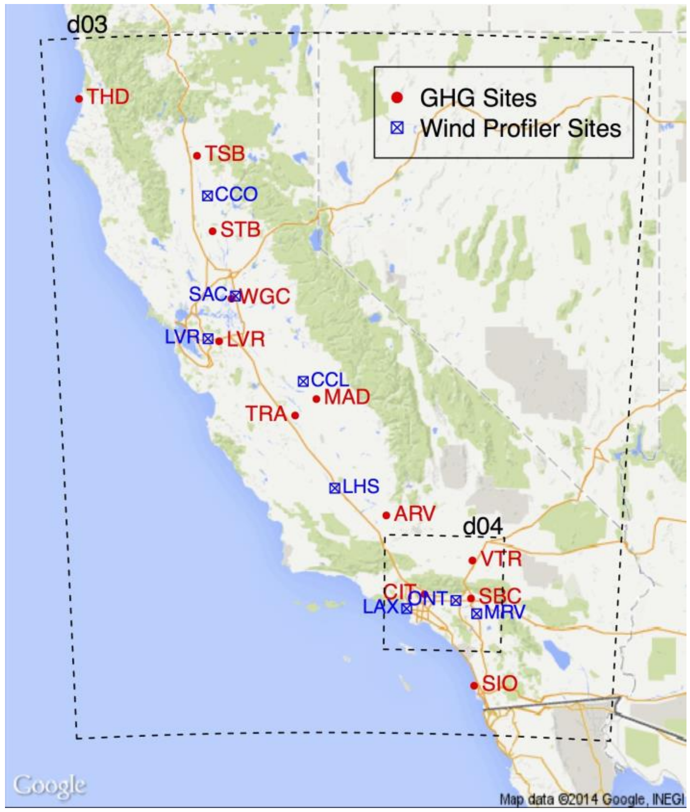

Figure 2. WRF simulation domains and locations of wind profiler sites used for the evaluation of WRF meteorology: CCO (Chico), SAC (Sacramento), LVR (Livermore), CCL (Chowchilla), LHS (Lost Hills), LAX (Los Angeles Airport), ONT (Ontario Airport) and MRV (Moreno Valley). Black dashed lines show the 4 km (d03) and 1.3 km (d04) domains for WRF simulations, covering California and SoCAB, respectively. The SFBA region is also simulated on the 1.3-km grid (not shown).

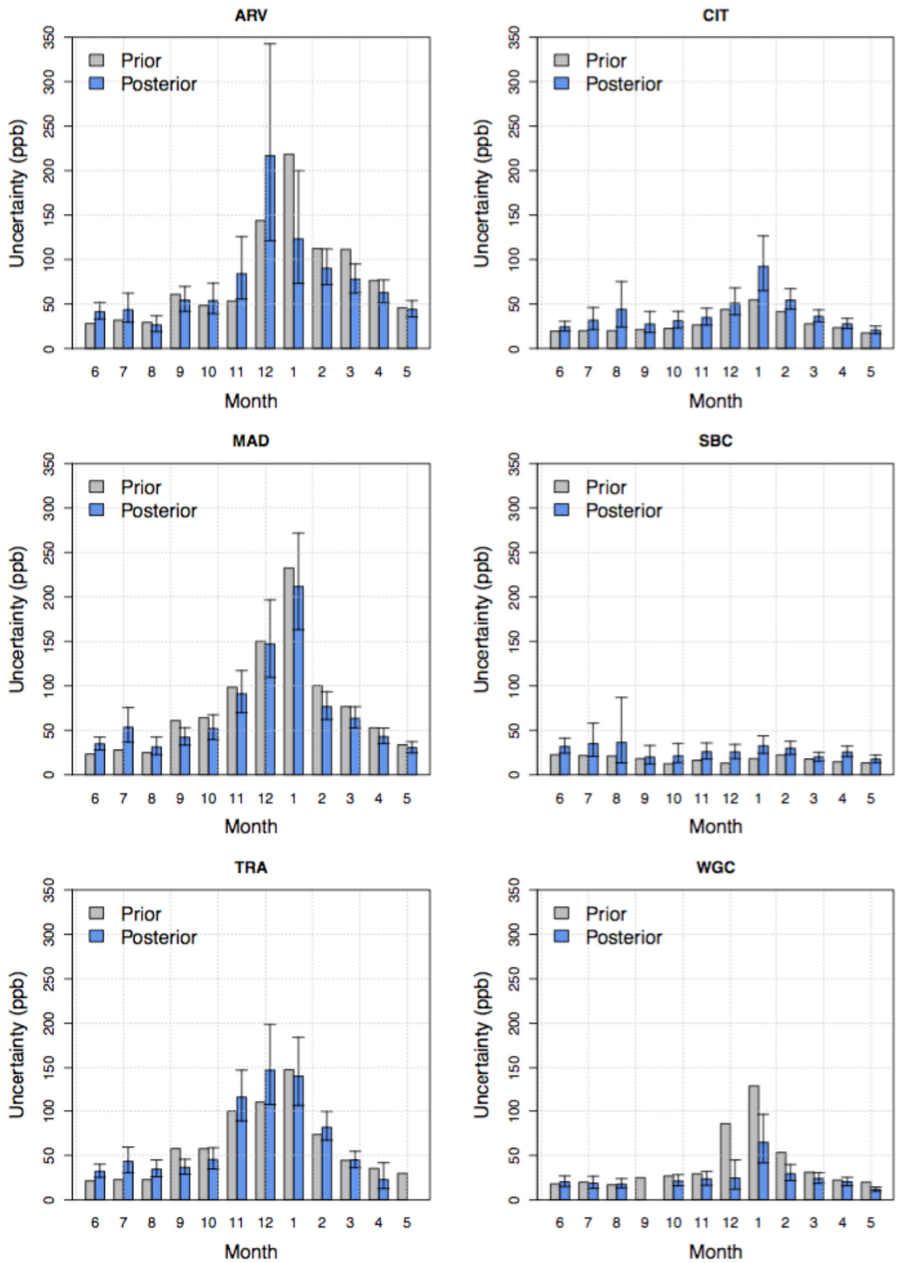

Figure 3. Estimated diagonal elements of the model-measurement mismatch matrix **R** for CH_4 inversions. The posterior values were estimated using 25000 MCMC samples and the error bar represents the 95% confidence interval. The prior values were estimated using the method described in Jeong et al. [2012a, 2012b, 2013] (see Text S1 in SI). For May at TRA and September at WGC, the posterior values were not estimated because most of the measurements were not available.

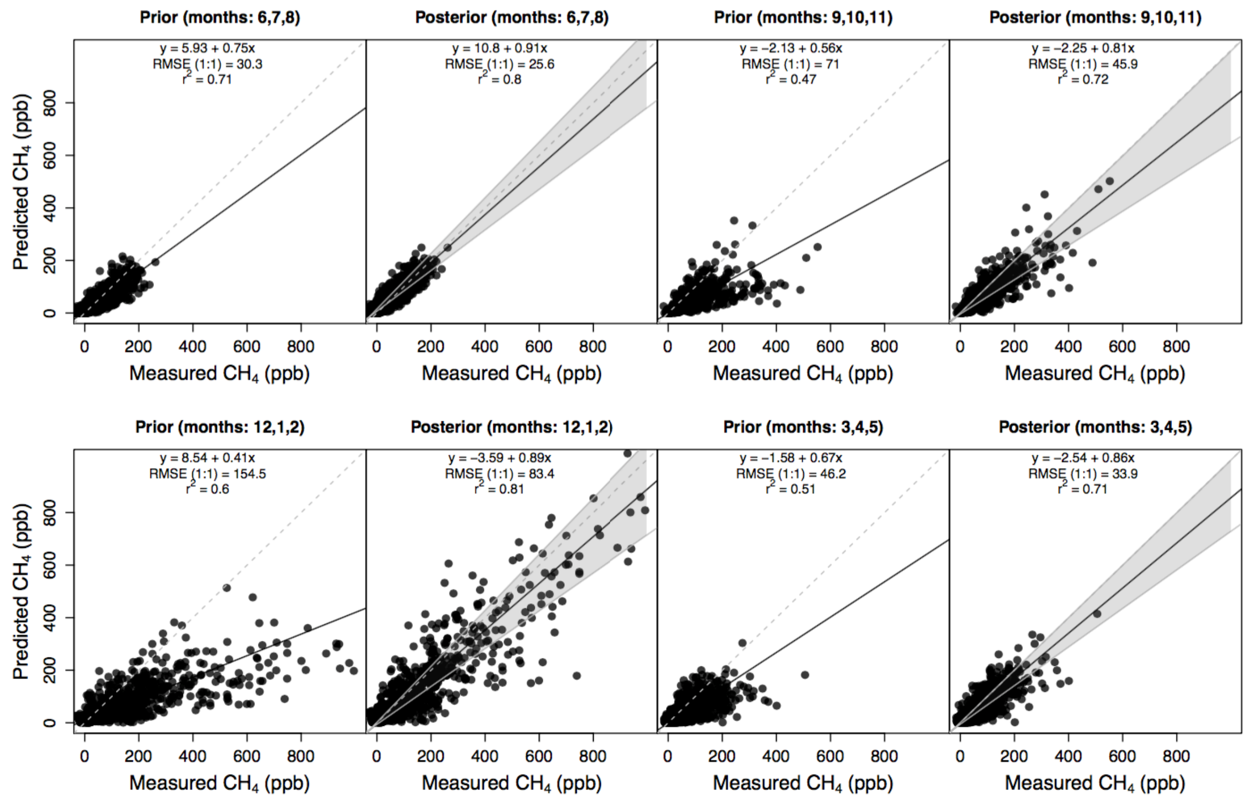

Figure 4. Comparison of predicted and measured CH_4 mixing ratios before (prior) and after (posterior) inversion for each season. The relatively low best-fit slopes in the prior comparison (left plot in each season) suggest prior emissions are underestimated. Filled circles represent individual 3-hour data points across different sites used in the inversion. The gray dashed line indicates the 1:1 line and the black solid line represents the best-fit slope for the data shown. The regression coefficients in the posterior plot were calculated based on the median values of the 25000 MCMC samples. The gray shaded area in the posterior plot represents the 95% uncertainty region for the regression analysis using 25000 MCMC samples.

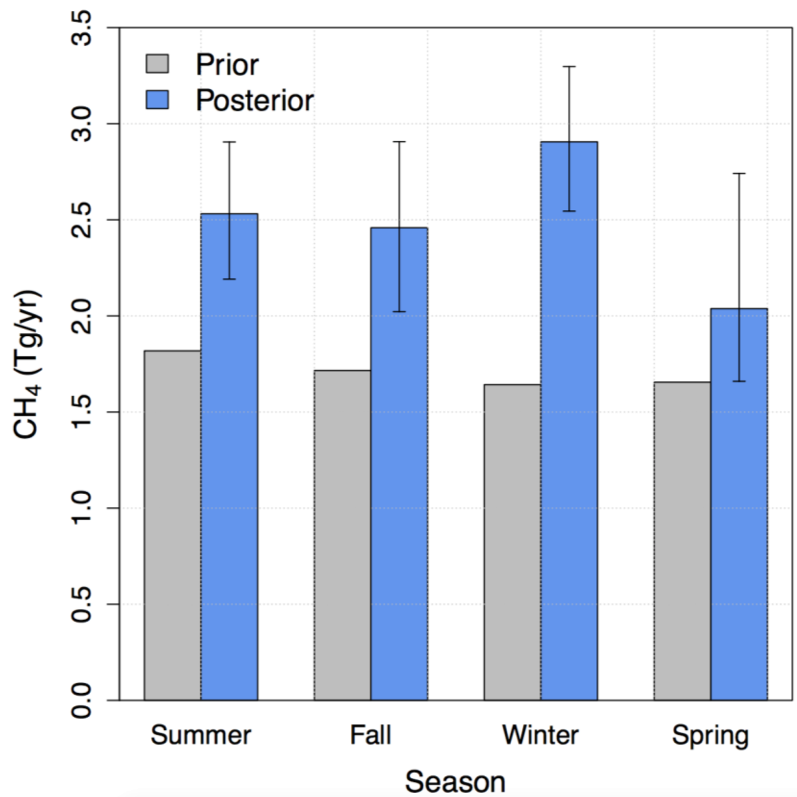

Figure 5. Inferred CH_4 emissions using measurements from 13 sites for four seasons: summer (JJA), fall (SON), winter (DJF) and spring (MAM). The error bar represents the 95% confidence interval around the median value of the posterior emission estimate.

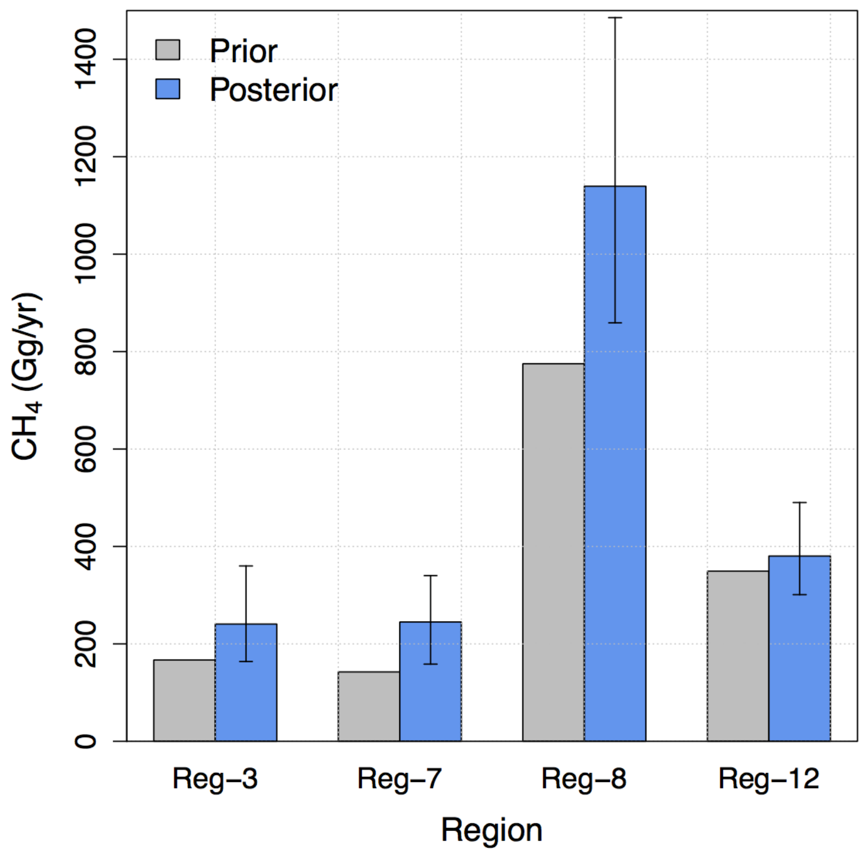

Figure 6. Estimated annual CH_4 emissions for the major emission regions (at 95% confidence). Regions 3, 7, 8 and 12 represents the Sacramento Valley (SV), San Francisco Bay Area (SFBA), San Joaquin Valley (SJV) and South Coast (SoCAB) air basins, respectively.

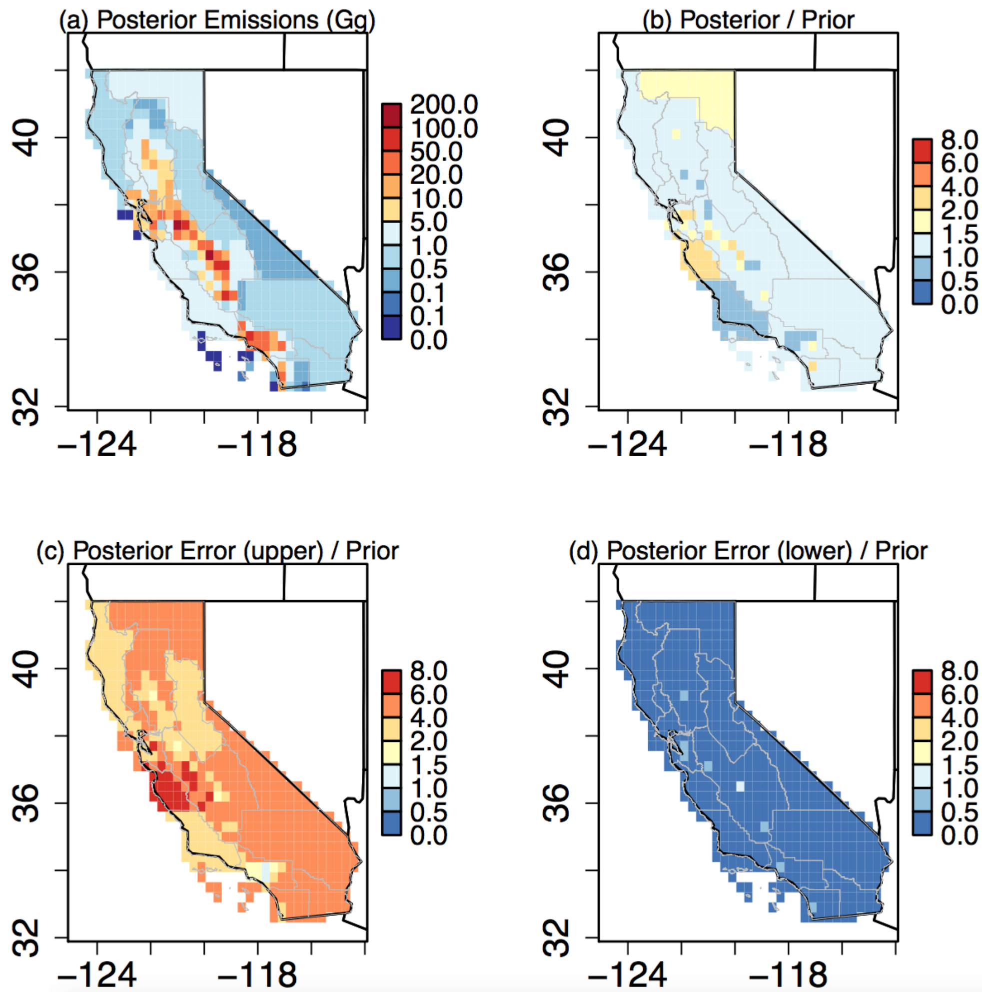

Figure 7. Estimated annual CH_4 emissions from the HBI analysis: (a) posterior (median) annual emissions (Gg/yr), (b) ratio of posterior to prior, (c) ratio of estimated 97.5th percentile to prior, and (d) ratio of estimated 2.5th percentile to prior.

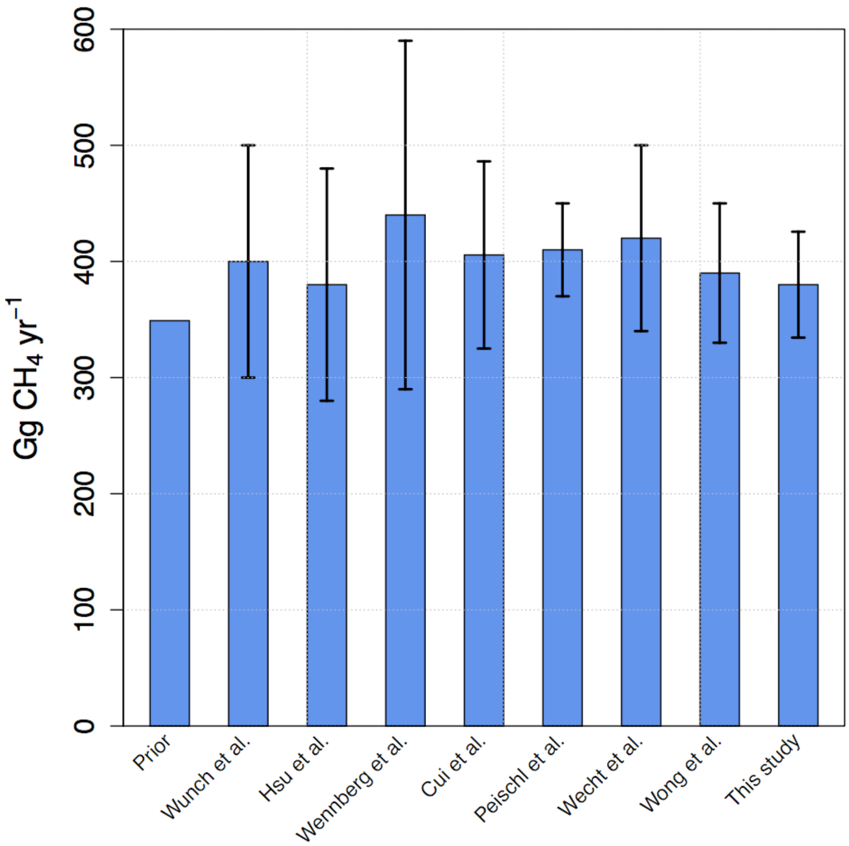

Figure 8. Comparison of the CALGEM prior (total for SoCAB = 349 Gg CH_4/yr) and estimated CH_4 emissions for SoCAB in the eight different recent studies including the posterior emission from this study. The value from Wunch et al. [2009] shows the CO-based estimate. Originally Hsu et al. reported LA County emissions (at 200 Gg CH_4/yr) and Wennberg et al. expanded the Hsu et al. results to the full SoCAB. The uncertainty estimates are 68% confidence intervals reported by the individual studies.

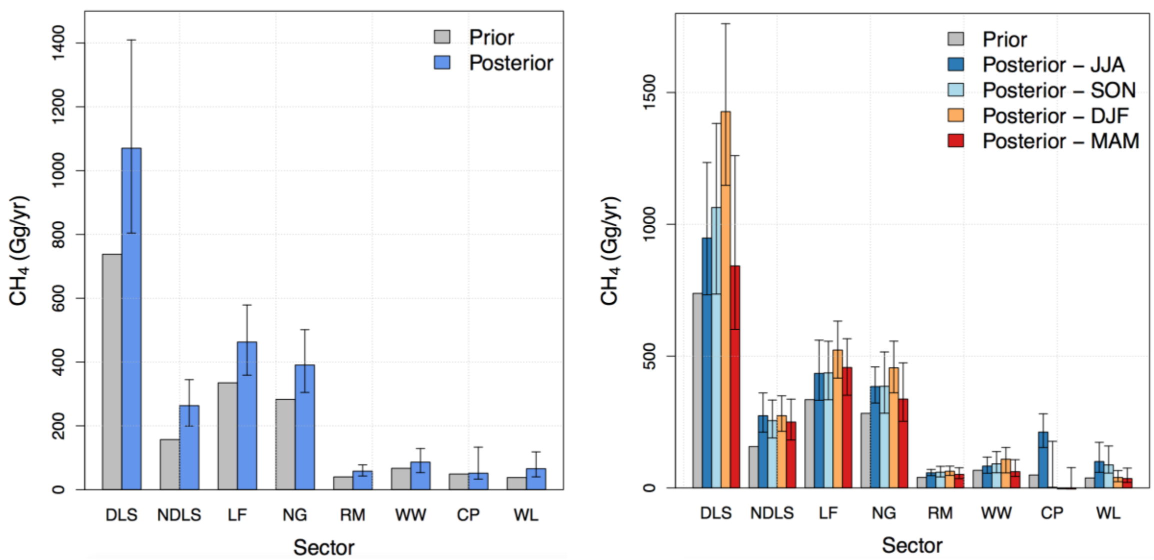

Figure 9. Posterior annual (left) and seasonal (right) emissions (Gg CH_4/yr) estimated from the HBI analysis by sector: dairy livestock (DLS), non-dairy livestock (NDLS), landfill (LF), natural gas including petroleum production (NG), petroleum refining and mobile sources (RM), wastewater (WW), crop agriculture (CP, largely rice), and wetland (WL). The error bar represents the 95% confidence interval.