Agriculture - Enteric Fermentation and Manure Management Emissions from Dairy and Beef Cattle Feedlots

Aaron Davitt[1], Gabriela Volpato[1], Krsna Raniga[1], Ann Marie Gardner[2], Lynn Henning[3] and Matthew Scheffer[2], Elliot Block[1], Akasha Sutherland[4], Brian Goodwin[4], Corey Jaskolski[4], Vivian Powell[5], Claire Pluard[5], and Sam Schiller[5]

1) WattTime, 2) Hudson Carbon, 3) Socially Responsible Agriculture Project, 4) Synthetaic, 5) Carbon Yield

1. Introduction

According to the Food and Agriculture Organization (FAO) data (FAOSTAT), beef and dairy milk production systems are the largest contributors of greenhouse gas (GHG) emissions in the livestock sector, representing more than 60% of emissions in the sector and 14.5% of all anthropogenic sources (FAO 2013). Beef and dairy sector emissions are driven by two sources. The primary source is enteric fermentation emissions which consists of methane (CH_4) gas produced in the digestive systems of ruminants and to a lesser extent non-ruminants. The secondary source is GHG emissions from manure management, producing both methane (CH_4) and nitrous oxide (N_2O) emissions via aerobic and anaerobic decomposition of livestock manure, including the microbially-driven processes of nitrification and denitrification (Waldrip et al., 2016; Waldrip et al., 2020). These emissions occur within manure storage facilities common to beef and dairy systems, as well as in-field where manure has been applied, or deposited by livestock.

The current de facto beef and cattle FAOSTAT emissions estimates are based on country-level self-reported activity data and constitute one of the most comprehensive global agricultural emissions datasets. While comprehensive at a global scale, such information is coarse with no specific facility attribution of emissions. Generally, facility level information is reported in academic studies, and, at times, the location information has been removed or kept vague (Harper et al., 2009; Costa et al., 2014; Zhu et al., 2014). However, while some jurisdictions have permitting databases that contain location data for livestock production data, the location of most livestock production facilities worldwide is unknown. As such, understanding facilities emissions and contribution to regional and global GHG emissions is a difficult process.

In order to understand individual beef and cattle facilities emission contributions, the Climate TRACE coalition has developed a method that makes novel use of an artificial intelligence tool called Rapid Automated Image Characterization (RAIC) and satellite imagery to identify beef and dairy feedlot facilities in California and Texas, U.S. and portions of Argentina. Climate TRACE then utilized facility level cattle population data from U.S. based permitting databases to build a cattle population prediction model. The basis of this model is the beef or dairy feedlot's area is related to the quantity of livestock confined. Once each facility's cattle population was estimated, IPCC emissions factors were applied based on the facilities' climatic conditions, production type, confinement area, and region, to predict facility level emissions for year 2021. These methods have generated a first of its kind database of facility level emissions estimates in the beef and dairy sector.

2. Materials and Methods

The approach utilized here primarily relies on the hypothesis that beef or dairy facility area size can be used as a predictor to estimate the total cattle population, which can then be used to estimate enteric fermentation and manure management emissions (Figure 7). Previous studies have shown the number of animals per farm, or the population, has a direct relationship to total GHG and non-GHG emissions (Harper et al., 2009; Vechi et al., 2022). This basis was applied to beef and dairy facilities in California and Texas, U.S. and portions of eastern Argentina for the year 2021. These regions were selected due to the high concentration of dairy and beef feedlots in their respective country. A combination of reported permitted beef and dairy populations, artificial intelligence, satellite imagery, modeling and the application of emission factors were used to estimate emissions for 2021 at each individual beef and dairy facility. The sections below provide a description of each dataset employed in this study.

2.1 Datasets employed

2.1.1 Feedlot population data

feedlot populations and locations from the following agencies:

1) **Animal waste regulated facility reports** from the California Environmental Protection Agency (EPA) State water resources control board (https://waterboards.ca.gov/water_issues/programs/ciwqs/publicreports.html). These reports provide detailed information including agency owner, street address, latitude and longitude of the feedlot facility and feedlot population. For the purposes here, beef and dairy feedlot facilities were filtered out from the reports using specific Standard Industrial Classification (SIC) codes 211, 212, 241 and 11212, identified as beef cattle feedlots, beef cattle except feedlots, dairy farms, and dairy cattle and milk production, respectively.

2) **CAFO wastewater general permits** from the Texas Commission on Environmental Quality (https://www2.tceq.texas.gov/wq_dpa/index.cfm). This permit data provided the information on wastewater discharge from specific facilities into or near adjacent water

bodies. The information here includes animal facility type, latitude and longitude of the facility, estimated waste solids generated, and feedlot population.

For this modeling effort, California and Texas beef and dairy feedlot population data was used, in addition to the latitude and longitude of each feedlot. We treat each state's population data as providing a representation of cattle numbers at specific feedlots reported in each database.

2.1.2 Remote sensing datasets

The following satellite imagery datasets were used to identify beef and dairy facilities:

National Agricultural Imagery Program (NAIP) aerial imagery distributed by The U.S. Geological Survey (USGS). NAIP acquires aerial imagery in the red, blue, green and near infrared wavelengths at a 1m spatial resolution, or finer, as part of an agricultural census conducted in the USA. Imagery was acquired every five years, starting in 2003, before switching to every three years starting in 2009. NAIP imagery acquired in 2020 were used to identify feedlot facilities in California and Texas with the RAIC tool (described in section 2.2 and 2.3.1). More information on NAIP imagery can be found at the USGS NAIP imagery program page (https://www.usgs.gov/centers/eros/science/usgs-eros-archive-aerial-photography-national-agriculture-imagery-program-naip).

To identify Argentinian beef and dairy facilities, **PlanetScope visual basemaps imagery created by Planet Lab's PlanetScope satellite constellation** was accessed for the RAIC tool (Planet, 2022). Each basemap was generated from optimal PlanetScope imagery comprising blue, green, red visual imagery. PlanetScope basemap was downloaded through Climate TRACE's API access. More information on Planet basemaps can be found on their website (https://developers.planet.com/docs/data/visual-basemaps/). For the purposes here, basemap imagery, from the 2020 summer months, covering central and northern Argentina were accessed and used with RAIC (described in section 2.2 and 2.3.1). Planet imagery was selected for this work due to the lack of global high spatial resolution satellite imagery similar to NAIP spatial resolution.

2.1.3 Remote sensing derived datasets

The Gridded Livestock of the World (GLW v3.1) dasymetric weighted gridded cattle density for 2010 (animals/km^2; Gilbert et al., 2018). The purpose of this dataset was to provide globally 10km gridded livestock distribution information for the year 2010, which includes other animal species aside from cattle. This dataset was an updated version of the initial GLW v1 dataset produced in 2007. The GLW v3.1 dasymetric weighted gridded dataset assigns cattle density to areas or boundaries based on environmental predictors. This is performed using a random forest model, where the cattle density distribution is based on different spatial predictors including land cover type, human population distribution, topography, vegetation classes and satellite derived vegetation indices, and climatic conditions (Gilbert et al 2018). This dataset was employed in our dairy modeling, described in further detail in section 2.4.1.

2.1.4 Temperature data

Average annual temperature at each individual feedlot facility from Meteosat (https://meteostat.net/en/). Meteosat provides global open source weather data from various agencies, including National Oceanic and Atmospheric Administration (NOAA), Deutscher Wetterdienst and Environment Canada. Meteosat provides a python package, allowing users to query specific latitude and longitude locations for temperature data. Here, average annual temperature data at each feedlot facility was identified and used to scale IPCC emission factors. For facilities with no associated temperature data, the closest feedlot facility with temperature data was used to fill in feedlots with missing temperature data.

The average annual temperature data was used to determine the IPCC "MANURE MANAGEMENT METHANE EMISSION FACTORS BY TEMPERATURE" in Table 10.14 (IPCC 2006).

2.1.5 IPCC emission factors

Emission factors from The IPCC Chapter 10: Emissions from Livestock and Manure Management and Chapter 11: N_2O Emissions from Managed Soils, and CO_2 Emissions from Lime and Urea Application (IPCC 2006a; IPCC 2006b). Dairy and beef emissions are produced two ways, 1) from enteric fermentation and 2) manure management. To estimate dairy and beef methane emissions from enteric fermentation and methane and N_2O emissions from manure management, Tier 1 approaches and a Tier 2 approach (for Indirect N_2O emissions due to leaching from manure management only) was applied along with default EFs based on region, temperature and manure management types, as outlined in IPCC guidelines. Assumptions were made for dairy and beef manure management- an uncovered anaerobic lagoon system for dairies (based on visual inspection of a subset of California and Texas facilities), and dry lot for beef (for all locations, based on IPCC definition). A list of emission factors is provided in Table 1. Direct and indirect N_2O emissions used default region Nitrogen excretion rate (Nex) values. The fraction of manure nitrogen that leaches from manure management systems ($Frac_{leachMS}$) ranges between 1 to 20%. For this work, $Frac_{leachMS}$ was set to 3% for all regions based on the reported total amount of nitrogen found in the soil profile from the excreted nitrogen from a feedlot in Harter et al. (2013).

2.2 Model Development - Rapid Automatic Image Categorization (RAIC) tool

Identification of beef and dairy facilities was performed in partnership with Synthetaic, a specialist in geospatial applications of machine learning tools including the Rapid Automated Image Characterization (RAIC) approach (https://www.synthetaic.com/). RAIC is a proprietary artificial intelligence tool developed by Synthetaic. This tool can be used to build AI models (Figure 1). By providing RAIC with raw, unlabeled imagery, this tool can be applied to large volumes of data to automatically find objects of interest. Additionally, using a human nudge tool, users can further refine the AI to better identify areas that contain a user-defined object of interest. From this, RAIC learns from the dataset to build an AI, which carries out an inference task to localize all cases of a user-defined object of interest, in this case feedlot facilities. For the work here, using a set of images of pre-validated livestock production facilities in each region, RAIC searched millions of square kilometers of remote sensing imagery and identified facilities with common features, identifying facilities that the tool was highly confident were confined to beef and dairy cattle operations.

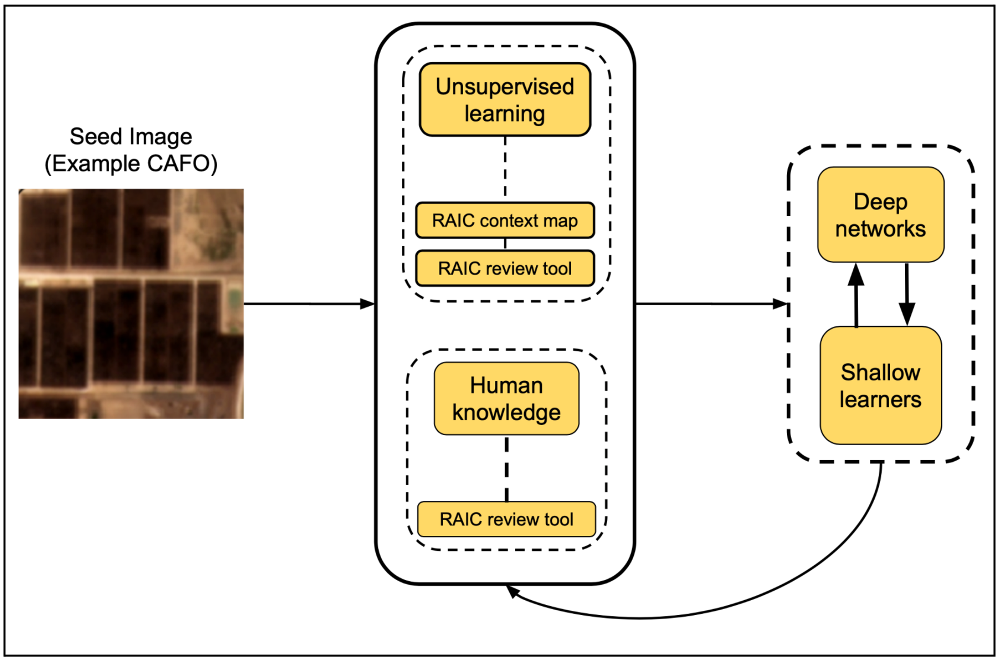

Figure 1 An example of the RAIC process to identify feedlots.

2.3 Methods

Figure 2 provides an overview of the model development and deployment to estimate beef and dairy feedlot area, then populations, which was then converted to Enteric Fermentation and Manure Management Emissions using IPCC emission factors. Below, each section describes the process leading into the emission estimates.

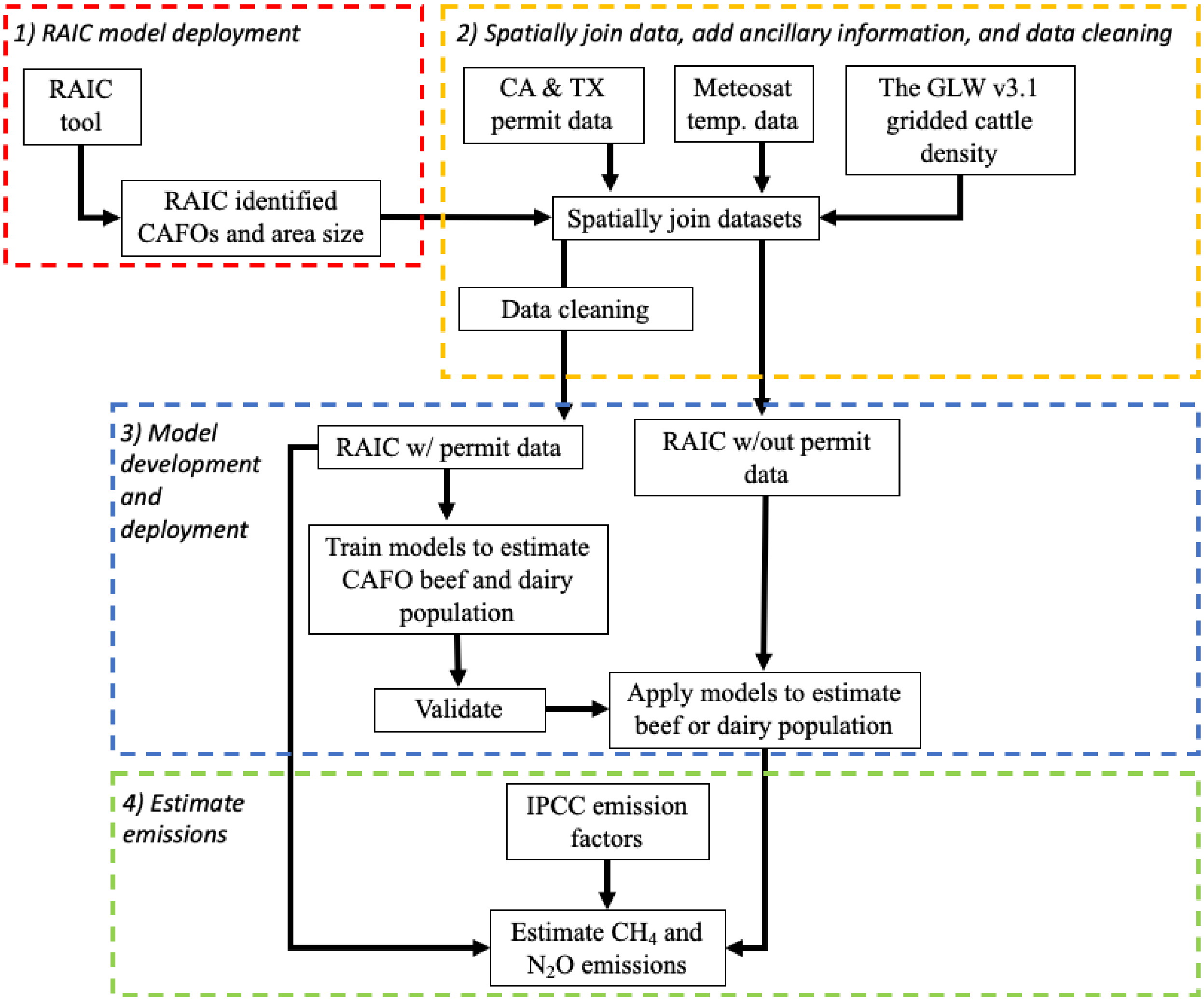

Figure 2 This flowchart summarizes the different section's inputs and outputs to estimate beef and dairy populations, and GHG emissions at individual feedlots. Starting at 1) *RAIC deployment*, the RAIC tool identified feedlots and produced facility boundaries in order to calculate their size. Supporting temperature, permitting, and regional data were added to facility locations in *2) Spatial join data, add ancillary information, and data cleaning*. Using these combined datasets, feedlot populations were estimated in *3) Model development and deployment*. Once the population estimates were derived, IPCC emission factors were applied in *4) Estimate emissions.* A description of each section is provided in the text.

2.3.1 RAIC model deployment

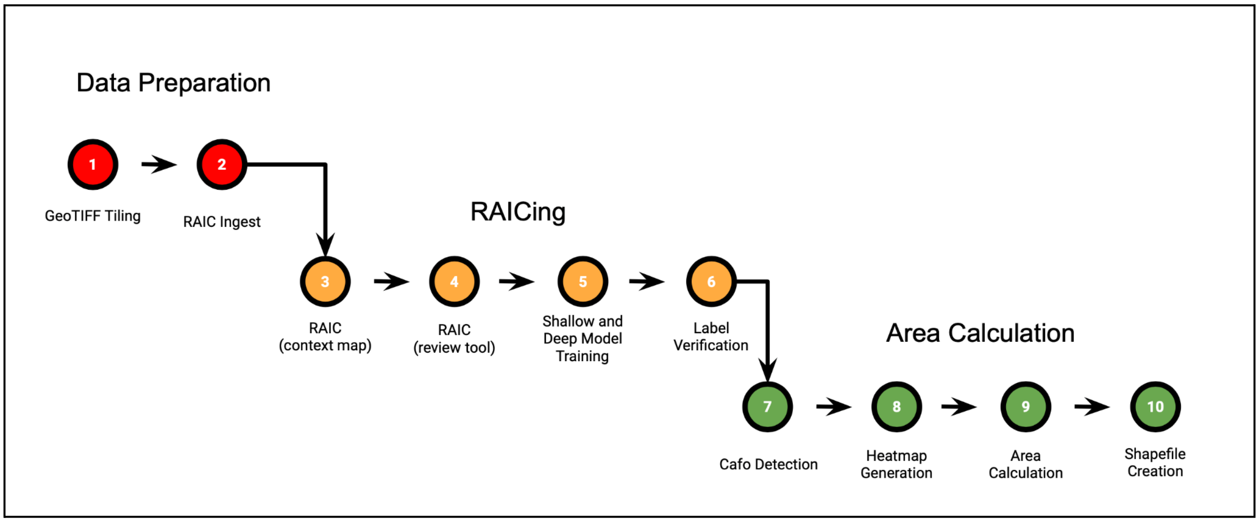

Figure 3 RAIC model deployment to identify feedlot facilities. Detailed description is included in the text.

The RAIC model was deployed to identify feedlot facilities in California, Texas and Argentina. This involved:

1) Data preparation - this includes downloading NAIP and PlanetScope basemap imagery and tiling (mosaicing) the imagery into the region of interest. Zoom levels were selected in which feedlot facilities would be identified. This includes using the California and Texas permit and Argentina data pulled from online sources to create a sample of beef and dairy feedlots for use as seed points in the RAIC model.
2) RAICing - described in section 2.2 and Figure 3. In short, a subset of California, Texas, and Argentina feedlot locations were provided to the RAIC tool as training data on what a feedlot looks like in the NAIP and PlanetScope basemap imagery. Then the RAIC tool learned to identify feedlots in other NAIP and PlanetScope imagery in each of these regions. Human knowledge helped filter what is and not a feedlot so the RAIC tool could learn and improve feedlot identification.
3) Area Calculation (Figure 4) - once a feedlot facility was identified, a convex hull heat map was generated to calculate the facility area. This is done by calculating the land mass inside each convex hull to generate a total area estimate for each identified feedlot. Coordinates were included for each area calculation.
4) Once completed, each feedlot identified was exported to a shapefile and spreadsheet that included area calculation and feedlot facility latitude and longitude centroid (Figure 5).

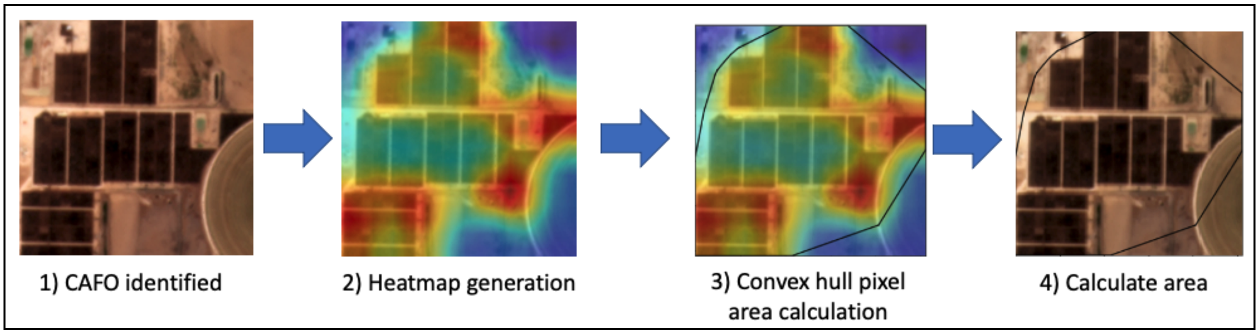

Figure 4 An example of the RAIC area calculation process.

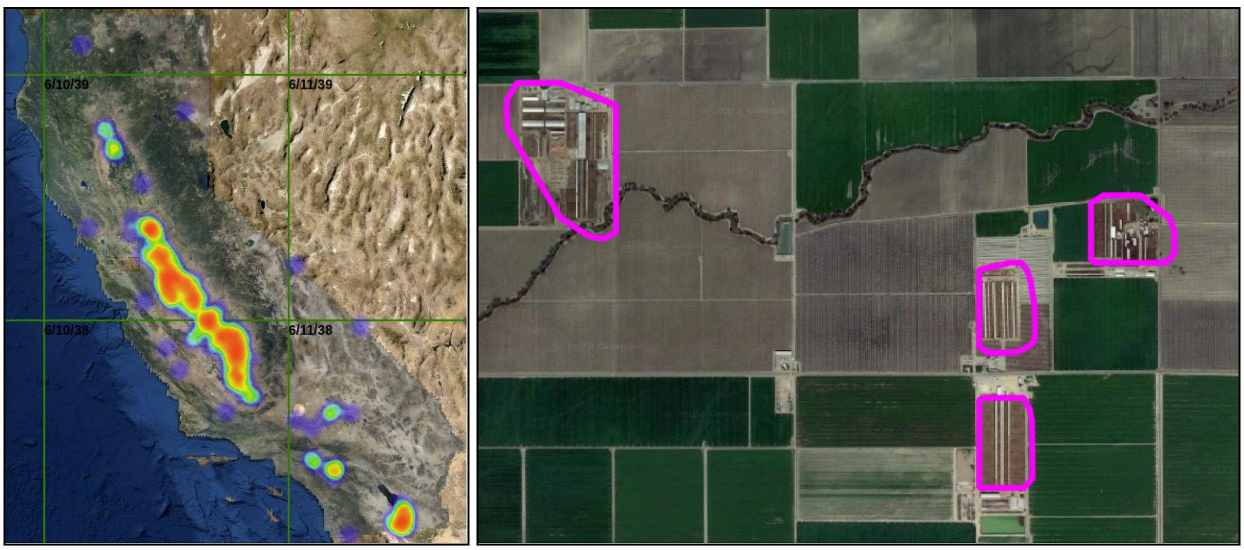

Figure 5 Left image, California heatmap of beef and dairy facilities. Red areas indicate higher concentration of beef and dairy facilities. Right image, example of RAIC identified facilities in California with area calculated.

2.3.2 ***Spatially join data, add ancillary information, and data cleaning***

To generate a beef and dairy population model for individual facilities required both an area (predictor) and reported population numbers (predictand). At the time of this research, no known dataset was available that contains both. Therefore, Climate TRACE created its own *in-situ* training dataset that contained both sets of information. As a first step, RAIC identified feedlots were spatially joined to California and Texas permit data that contained population information. This created 'one-to-one' matches (feedlots in dashed green boxes in Figure 6).

However, to ensure there was a large enough sample size for model training, another step was performed that manually geolocated California and Texas permit data to RAIC identified areas. In most cases with the spatial join, permit data locations fell outside RAIC areas (feedlot in green square in Figure 6). These unmatched facilities required manually aligning permit locations to fall within each RAIC area by using geographic information system (GIS) software. Some facilities required using Google maps and street view and/or searching for the business address to confirm the permit locations were matched to the correct areas. These 'aligned one-to-one' matches were added to the *in-situ* training dataset for model development. As a final step, temperature and gridded cattle density from Gilbert et al. (2018) was added to each individual facility in the training dataset.

In total, the spatial matching and aligning of locations created a model training dataset that contained both RAIC areas and population information. In total, 617 *in-situ* feedlots were created for model development- 36 in Texas (18 beef and dairy facilities each) and 581 in California (98 beef and 483 dairy facilities; Figure 7).

There were some RAIC identified feedlot areas that were absent from permitting data (unmatched areas; red circle in Figure 6). This includes all Argentinian RAIC feedlots. Additionally, there was California permit data that was not joined to any known RAIC areas (unmatched permits; red dashed circle in Figure 6). For both, the temperature data and gridded cattle density from Gilbert et al. (2018) were added to each unmatched set.

In summary, the following dataset were created and used as described:

1) A training dataset, containing matched feedlot areas to registered permit population data. This was used to develop the population model.
2) A dataset of unmatched areas (containing RAIC areas only). This dataset had the population model applied to predict the cattle population.
3) A dataset of unmatched permits (containing California permit population data only). This dataset was not used in the modeling as it already has population data.

As part of the process described above, data cleaning was performed to remove mislabeled facilities and provide accurate geolocation information. The following steps were performed to remove errors from the dataset:

1) California permit data did report cattle counts of less than 100 at facilities. To confirm if these locations had counts less than 100, a subset was selected for visual inspection in Google Earth Pro where each beef or dairy cow was counted at each feedlot. Based on this subset, it was determined that these locations had more than 100 head of cattle, a discrepancy in the permit data. As such, any California permit data with reported cattle population less than 100 was excluded from the model training dataset. We treated these feedlots as locations without population data.
2) California permit data had some mislabeled permit locations that could not be attributed to any identifiable facility. Some of these mislabeled locations were in the middle of crop fields or cities. At times, permit data could not be matched to a RAIC area and these were removed if no nearby facility was identified or could not be manually identified on

Google maps and searching for the business address. These permitted facilities were entirely removed.

3) In some cases, RAIC misidentified areas as large feedlots. For example, an airport or bare area. These RAIC areas were entirely removed.
4) The RAIC tools sometimes aggregated multiple feedlots into one large facility. If there were multiple permits associated with one identified feedlot, these were excluded from the model to avoid adding bias to the model training data. In these cases, the permit data was not associated with an RAIC area and added to the dataset of unmatched permits.
5) For Argentina, feedlots were visually inspected to remove any false-positives that may have been misidentified as beef or dairy feedlots. Misidentified locations were removed entirely.

This data cleaning was performed for all regions. However, mislabeled facilities may exist in the dairy and beef feedlot dataset on the Climate TRACE website, and we welcome users to report these facilities if identified.

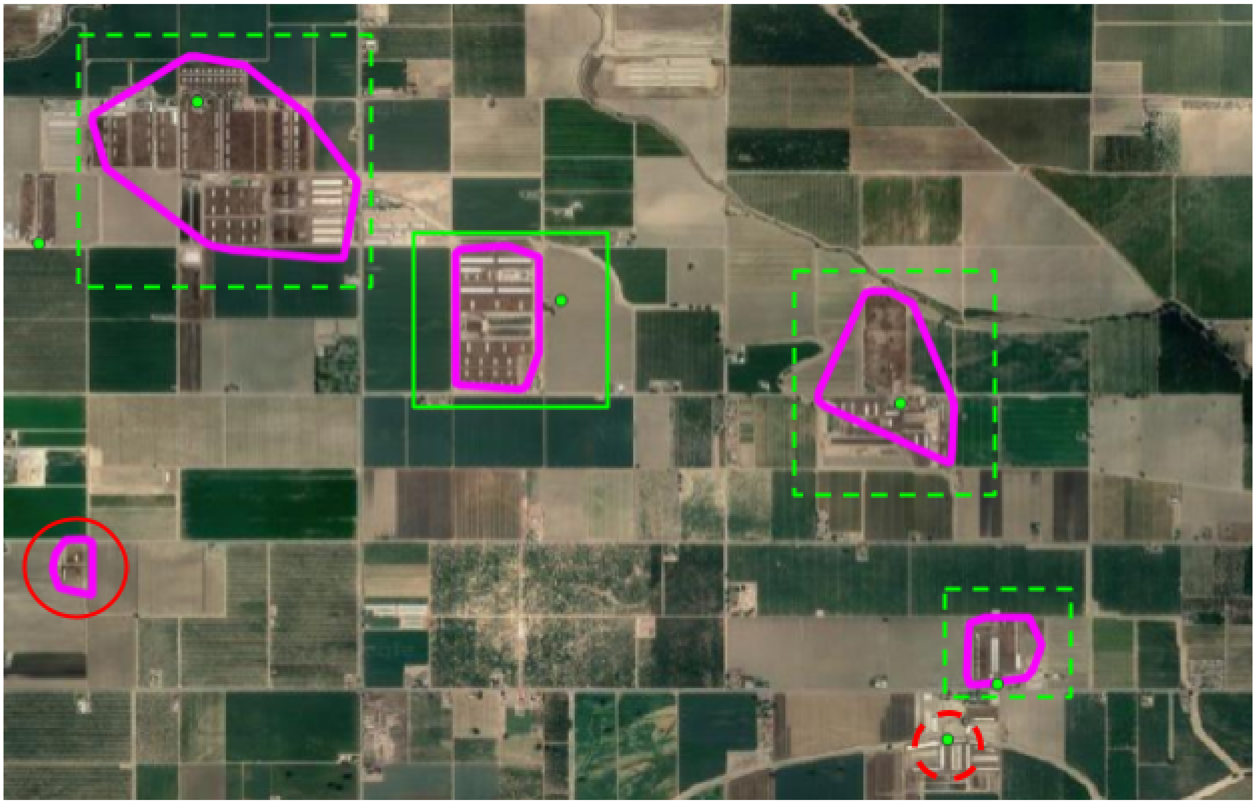

Figure 6 Example of matching permit data to RAIC area estimates. Green dots are California permit data and purple area shapes are RAIC identified feedlots. Green dashed boxes represent ‘one-to-one’ matches- permit data matched the RAIC area; the red circle identifies a feedlot facility identified by the RAIC tool but with no known permit data; the green rectangle represents a permit location outside the RAIC area boundary, an ‘aligned one-to-one’ example; and the red dashed circle (lower right) indicates a California permit data that was not matched to a RAIC area.

2.4 Modeling cattle populations at each facility

2.4.1 Population model development

The model development primarily relied on the relationship between beef and dairy feedlot size to the total cattle population. Using the training dataset, our analysis found that the feedlot area and the number of cattle at a facility were related (Figure 7). Based on the R^2 values and statistical significance ($p<.001$) shown in Figure 7, this relationship supports our hypothesis that feedlot area can be a predictor of cattle population, which then can be related to the total emissions estimate at individual facilities.

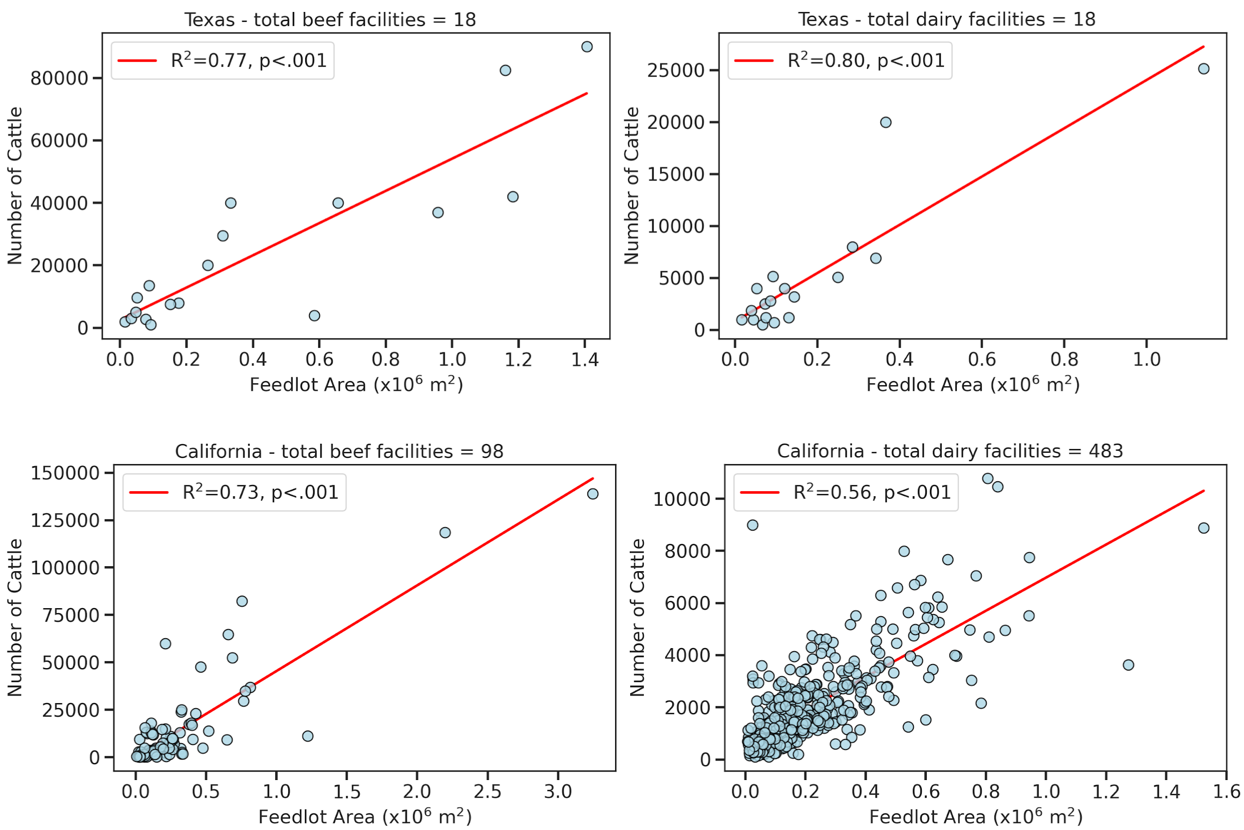

Figure 7 Relationship between feedlot area size and number of cattle for Texas beef (top left) and dairy (top right), and California beef (lower left) and dairy (lower right) facilities. R^2 values and statistical significance are included for each region by facility type. Note, y-axis range differs for each plot.

The training dataset developed in section 2.3.2 was split into beef and dairy sub-datasets for separate model development. For each model, a train-test-split was performed. For each sub-dataset, 80% of the data was used for training and 20% was set aside for testing through random sampling. Several different versions of the model, including random forest and gradient boosting regressor, and linear regression were tested. It was found that linear regression performed better relative to the other two models based on optimizing for mean square error (MSE). Results of this tuning are shown in the Results section.

Beef model

To estimate beef cattle population at individual beef facilities, a linear regression model was developed:

$$Beef\ cattle\ pop._i = f_x(Feedlot\ area\ km_i^2)\ (Eq.1)$$

Where 'i' is the individual beef feedlot being modeled.

Dairy model

To estimate dairy cattle population at individual dairy feedlot, the following linear regression model was developed that incorporated additional variables:

$$Dairy\ cattle\ pop._{i} = f_{x}(Feedlot\ area\ km^{2}_{i}, [Feedlot\ area\ km^{2}_{i}]^{2}, Density\ ratio_{i})\ (Eq. 2)$$

Where, "i" is the individual dairy feedlot being modeled, $[Feedlot\ area\ km^{2}]^{2}$ is the square of $Feedlot\ area\ size\ km^{2}$ and $Density\ ratio,$ which is defined as:

$Density\ ratio_{i,r} = Feedlot\ area\ ratio_{i,r} \times Ct\_2010\_Da_{i,r} (Eq. 3)$ and,

$Feedlot\ area\ ratio_{i,r} = Feedlot_{i,r}\ area\ (km^{2}) / \Sigma\ Feedlot\ area_{r}\ (km^{2})\ (Eq. 4)$

Where 'r' is a specific region (California, Texas, or Argentina) and 'Ct_2010_Da' is The Gridded Livestock of the World (GLW v3.1) dasymetric weighted gridded cattle density for 2010 (animals/km^2) from Gilbert et al. (2018). Instead of assigning the gridded cattle density value directly to an individual facility, a percentage was assigned, using the $Feedlot\ area\ ratio$, which is the individual Feedlot area relative to the sum of all feedlot areas identified by RAIC in a region. The $Density\ ratio$ was included as we found it improved model performance by accounting for dairy cow density in real, on the ground examples. Based on our estimates pulled from academic reports describing facility design, the stall width and length requirements of dairy facilities tended to allocate less space to dairy cows, ~2.8m^2 to ~3.2m^2, compared to beef cows, ~2.3m^2 to ~14m^2 (Euken et al., 2015; McFarland and Tyson 2016; Krekelberg 2020). With this, we hypothesized that assigning dairy facilities a $Density\ ratio$ helped to account for the slightly higher dairy cattle density at facilities. Including this variable helped improve model performance.

2.4.2 Model deployment

Beef and dairy feedlots with no population data were modeled using Eqs. 1 and 2, respectively. Not all Texas and California feedlots were identified as beef or dairy. Because permitted beef facilities were significantly larger than permitted dairy facilities in California and Texas, a size threshold was set based on the training dataset to categorize facilities. Any Texas or California feedlot area less than the maximum dairy area for both regions was tagged as dairy and feedlots greater than the maximum dairy size were tagged as beef. For Argentina, visual inspection was performed to tag beef or dairy feedlots.

Beef and dairy facilities with permit population data previously used to train the model were not modeled again as they already contain documented population information.

2.4.3 Estimate emissions

IPCC equations and default regional emission factors were used to estimate enteric fermentation methane, manure management methane, and nitrous oxide emissions (Table S1 in supplementary section). Temperature data at each location from 2021 adjusted the emission factors for specific facility regions. Visual inspection of feedlots was performed to determine the manure management type and emission factors, which was assumed to be constant for all feedlots in a region. The exception was California and Texas dairy manure management. We identified anaerobic lagoons, which IPCC assigns a 0 emission factor. However, studies by Owen and Silver (2011) and Petersen (2018) indicate that current manure management emission factors may underestimate emissions. As such, we apply the liquid/slurry emission factor of 0.005 to this feedlot type.

In total, emission estimates for 2021 were provided for 2388 feedlots:

- Argentina- 146 identified, 138 beef and 8 dairy feedlots.
- Texas- 666 identified, 468 beef and 198 dairy feedlots.
- California- 1576 identified, 278 beef and 1298 dairy feedlots.

On the Climate TRACE website, enteric fermentation and manure management were reported separately for each feedlot type. The reported data has the following information:

1) The feedlot area is defined as 'capacity', reported in km^2, which limits the number of cattle at each facility. For non-modeled facilities that already have permit population data but not an area, the capacity was set to '1', assuming that these feedlots have as many animals that it can fit. For modeled feedlots, we provide RAIC area km^2.
2) The estimated animal population of the feedlot is provided under the 'activity' column.
3) Manure management provides methane and nitrous oxide emissions separately and combines them into total CO_2 equivalent (CO_2e) for 20 year and 100 year global warming potentials (GWPs), reported as 'total_CO2e_20yrGWP' and 'total_CO2e_100yrGWP'. To convert methane to CO_2e, we applied a 80.8 (20 year) and 27.2 (100 year). To convert nitrous oxide to CO_2e, we applied 273 for each unit of N_2O.
4) Manure management provides information on the nitrous oxide emission factors in the 'other' columns.
5) The N_2O emission factors were created by dividing the N_2O emissions column by the activity column. The N_2O emissions factor combines the N2O emissions factor for direct emissions, N_2O emissions factor for indirect volatilization, and N_2O emissions factor for indirect leaching.

2.3.4 Verification of approach

To evaluate the beef and dairy model's accuracy to predict cattle population at individual facilities, 20% of the beef and dairy in situ datasets was set aside for this purpose. The metrics used to determine model performance were based on minimizing the mean square error (MSE) and the goodness-of-fit measure (R^2) for linear regression. Additionally, root mean square error (RMSE), mean absolute error (MAE), and Spearman's rank correlation coefficient (Rs) were included.

3. Results

3.1 Modeling results

Figures 8 and 9 report the model performance relative to facility reported permitting data that was withheld from data used to train the population model. A Spearman's rank correlation coefficient (Rs) is provided to assess the relationship strength between permit and modeled beef population. The beef Rs value is 0.70 ($p<.001$), displaying a strong relationship between the two, indicating the model predictor (area size) displays a relationship to the population size at a beef feedlot. The beef model is able to capture the variation observed in population sizes at different facilities. Of the samples, only four show disagreement, which occurs where beef populations are generally greater than 20,000 (Figure 8).

For the dairy model performance, the dairy Rs value is 0.73 ($p<.001$), displaying a strong relationship between the two (Figure 9). This suggests the dairy model predictors display a relationship to the population size at a dairy feedlot. The dairy model is able to capture changes in population numbers of different sizes. Disagreements do occur, generally when the dairy population is greater than ~4,500-5000. This suggests that facilities of a certain size may have a density factor that needs to be further accounted for, discussed in section 2.4.1.

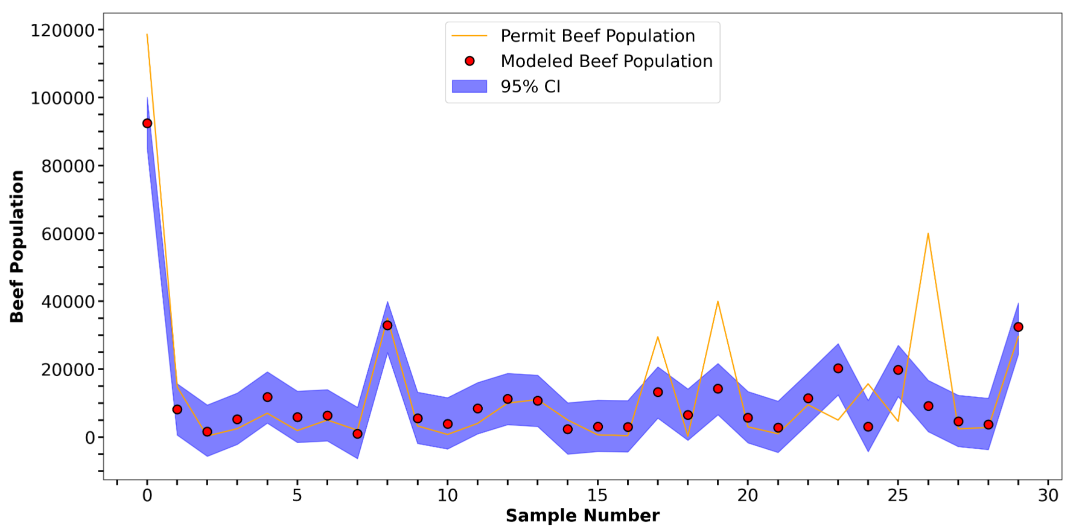

Figure 8 Beef modeling prediction interval results. Orange line is Permit Beef Population, red circles are Modeled Beef Population, and shaded area is the 95% confidence interval.

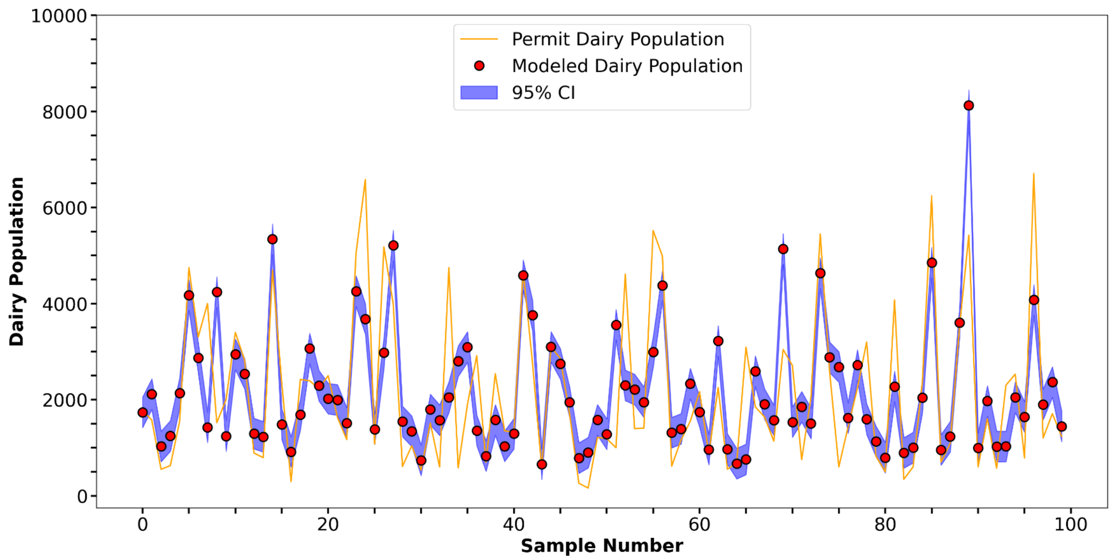

Figure 9 Dairy modeling prediction interval results. Orange line is Permit Dairy Population, red circles are Modeled Dairy Population, and shaded area is the 95% confidence interval.

The beef MSE and R^2 are 168,095,873 and 0.71 ($p<.001$), respectively (Figure 10). Generally, for beef feedlots with populations less than 20,000, the predicted values fall much closer to on the 1:1 line. Beef feedlots with a population greater than 20,000 tend to be underpredict relative to the reported beef population. This underprediction could be attributed to fewer training samples with populations greater than 60,000. For dairy, the MSE and R^2 are 1,244,516 and 0.51 ($p<.001$), respectively (Figure 10). Generally, for dairy feedlot less than ~2,000, the predicted values fall much closer to the 1:1 line. Feedlots with a population greater than ~3,000 tend to have underestimated populations relative to the reported population. As with beef, this underestimate could be attributed to the relatively fewer facilities with populations greater than ~300 in the training data.

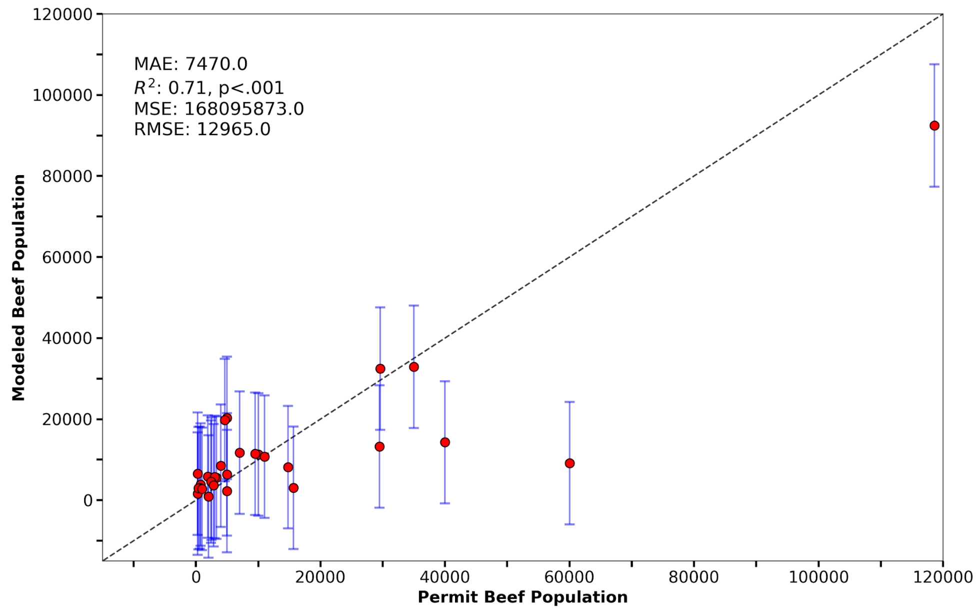

Figure 10 Scatter plots comparing reported beef population (x-axis) to estimated (modeled) beef populations (y-axis). Blue vertical lines are 95% confidence intervals.

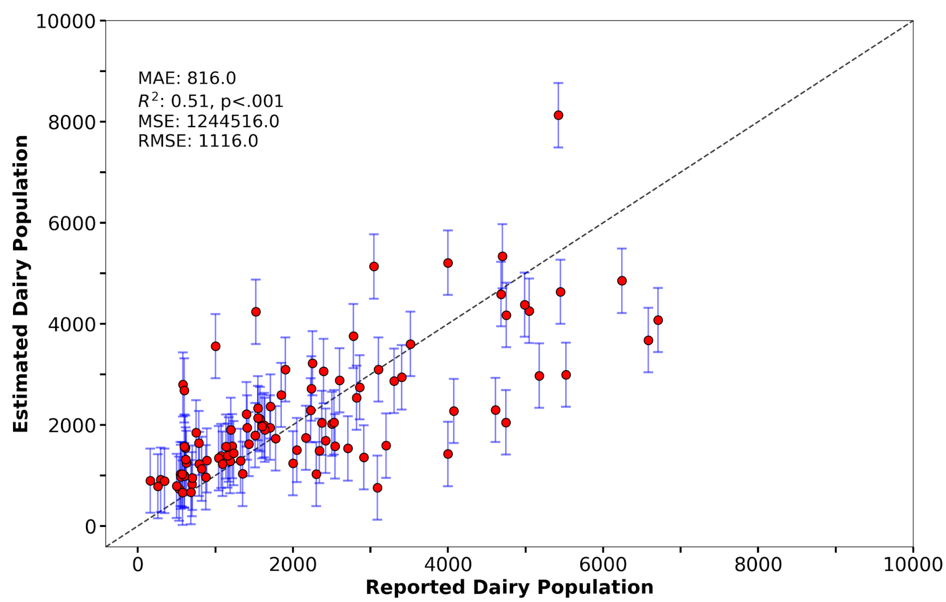

Figure 10 Scatter plots comparing reported dairy population (x-axis) to estimated (modeled) dairy populations (y-axis). Blue vertical lines are 95% confidence intervals.

3.2 Emission estimates

Regional emission estimates are provided in Figures 11 to 13, with Table 1 providing minimum, mean, and maximum emissions values by feedlot type and source. Based on the feedlot facilities identified in our approach, Texas beef enteric emissions are the largest of the three regions, a mean value of 53,469 metric tonnes. In California, this is 27,962 metric tonnes and 33,676 metric tonnes for the area surveyed in Argentina. Overall, beef feedlots in these regions have significantly larger populations relative to dairy feedlots, resulting in higher emissions. However, from a manure management perspective, dairy feedlots have larger emissions relative to their beef counterparts. Texas has the largest dairy manure management emissions, a mean of 12,197 metric tonnes, where California and Argentina have mean values of 8,257 and 85, respectively. This is due to dairy having different manure management practices compared to beef, which produces more emissions. Argentina's manure management values are significantly lower than Texas and California. This may simply reflect that the region of Argentina that was RAICed is a more heavily beef producing region or that RAIC was not given sufficient seed points to train detection of dairy facilities.

The majority of feedlots emissions in California are concentrated in the Central Valley, where significant agricultural production occurs. In Texas, the emissions are concentrated in the northern panhandle region. In Argentina, the feedlots identified by RAIC are primarily concentrated in the Pampas, in the central part of the country.

Table 1 Minimum, mean, and maximum emissions estimates by type and region. Values shown are Total CO_2e 20 year GWP reported in metric tonnes.

	Argentina				**California**				**Texas**			
	Enteric fermentation		*Manure management*		*Enteric fermentation*		*Manure management*		*Enteric fermentation*		*Manure management*	
	Beef	Dairy	Beef	Dairy	Beef	Dairy	Beef	Dairy	Beef	Dairy	Beef	Dairy
Min	2,941	3,572	53	50	428	1,034	11	412	1,794	5,171	42	2,835
Mean	33,676	6,141	601	85	27,962	14,612	1,055	8,257	53,469	22,851	1,743	12,197
Max	387,223	13,066	6,915	181	594,787	111,450	22,524	61,822	705,834	259,842	18,028	150,228

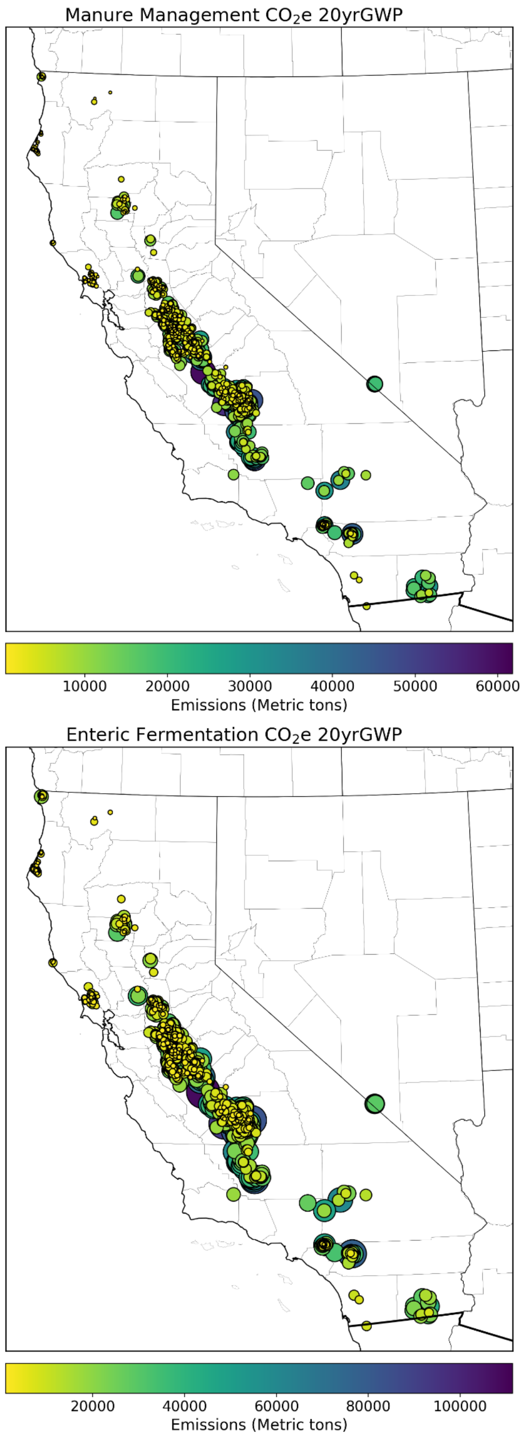

Figure 11 California dairy emissions (left column) and California beef emissions (right column) by emissions type: manure management (top row) and enteric fermentation (bottom row). Emission values are metric tonnes reported as 20 year Global Warming Potential (CO_2e 20yrGWP).

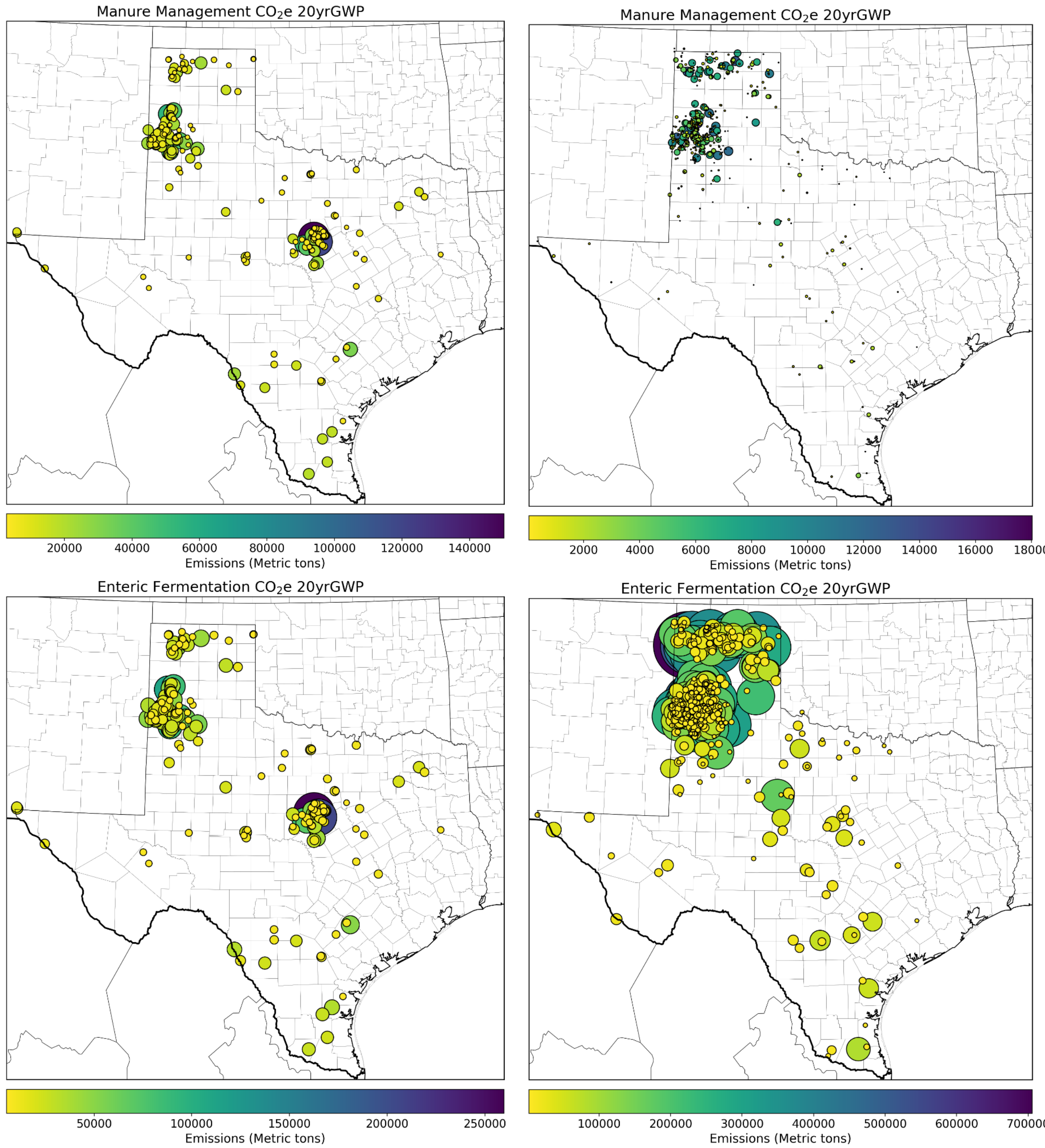

Figure 12 Texas dairy emissions (left column) and Texas beef emissions right column) by emissions type: manure management (top row) and enteric fermentation (bottom row). Emission values are metric tonnes reported as 20 year Global Warming Potential (CO_2e 20yrGWP).

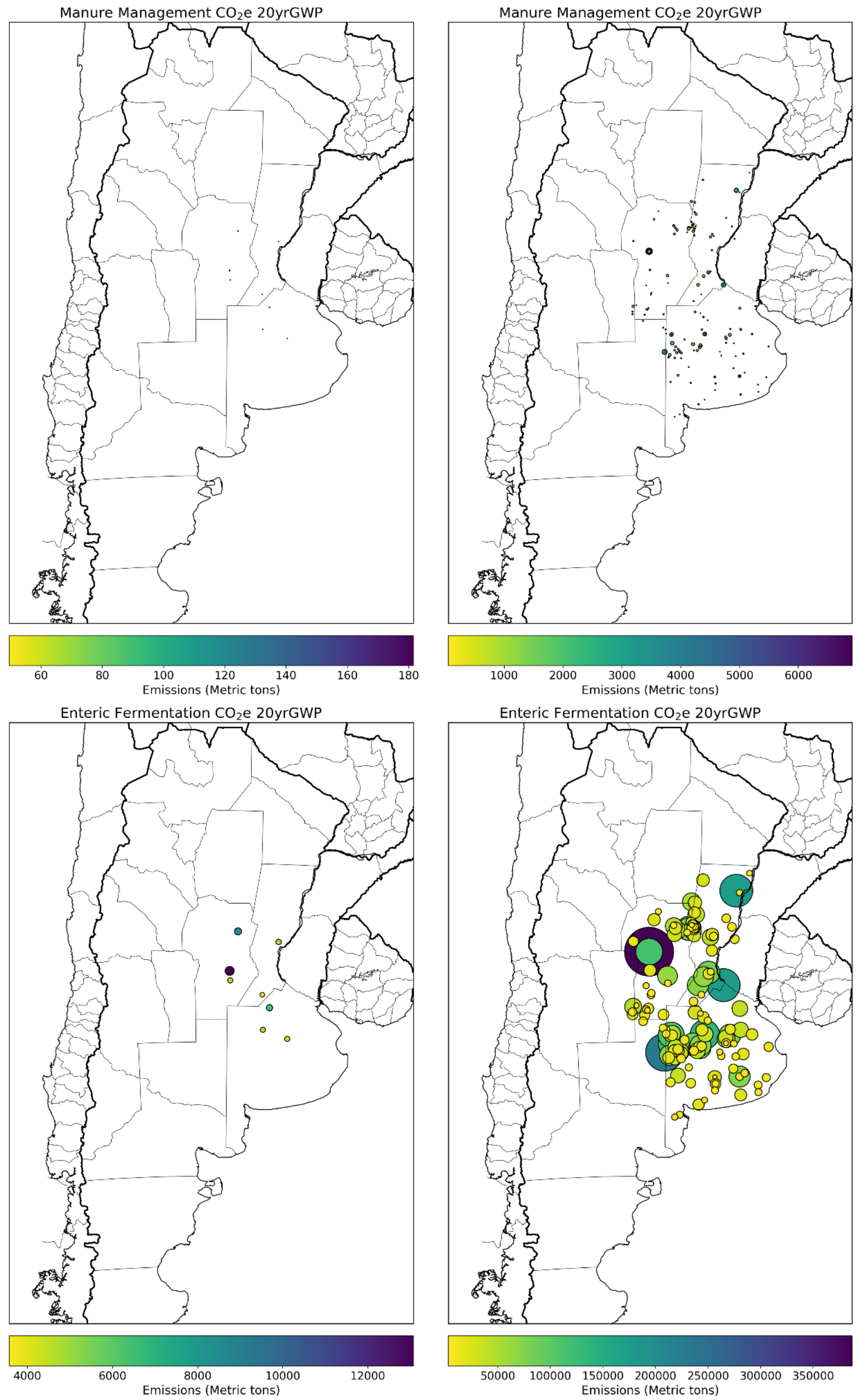

Figure 13 Argentina dairy emissions (left column) and Argentina beef emissions right column) by emissions type: manure management (top row) and enteric fermentation (bottom row). Emission values are metric tonnes reported as 20 year Global Warming Potential (CO_2e 20yrGWP).

4. Discussion and Conclusions

Climate TRACE's deployment of the RAIC tool represents a scalable approach for spatially mapping and estimating facility level emissions from beef and dairy feedlots. This approach leverages available permitting data, temperature, and gridded cattle density to train a model to estimate livestock population for facilities without publicly available population data. These paired approaches of facility identification and population modeling, alongside the application of relevant emissions factors, represents one of the first attempts to downscale emissions modeling to the facility level across an entire state or province.

There are several conditions where this modeling approach works well for facility level emissions modeling. RAIC is best at identifying facilities in contexts where beef and dairy feedlots are sufficiently large and differentiated from nearby land uses. As a rule of thumb, RAIC is only capable of correctly labeling facilities in contexts where the human eye can successfully differentiate a feedlot from nearby land uses. Thus, small facilities that may blend into the landscape, areas with low resolution imagery or high levels of haze that obscure aerial photography, extensive grazing lands that look similar to crop lands, or indoor facilities near adjacent enclosed buildings may be missed when utilizing the RAIC tool.

The Climate TRACE modeling approach works best when there are ground-truthed, facility-level population counts at a representative portion of feedlots, and where manure management is relatively standardized in a region. These conditions are not present in all beef and dairy production regions, which may limit the applicability and precision of this approach. The population modeling requires some degree of local tuning to ensure that regional differences in management do not impact the cattle density feedlots. It was infeasible at this phase of the research to label the manure management practices of every facility in the dataset or indicate where facilities may have had mixed beef and dairy production systems. As such, we assumed the same manure management practices for feedlots by type in each region. Facilities were labeled either beef or dairy facilities, blurring the differences between integrated facilities. Differences in manure management between feedlots in a region were also difficult to capture in these results. All beef facilities in an area were assigned one regional manure management practice (as were all dairy facilities) that was observed to be most common to that geography. The model also did not attempt to adjust facility emissions to account for differences in utilization. Permit records used to estimate feedlot population indicate a maximum capacity that may not be fully utilized year round. To the extent that there are differences in utilization within a region, the results may overestimate livestock emissions.

Future modeling will attempt to address several limitations of this process. Remote sensing approaches that detect features like manure lagoons, slurry pits, or anaerobic digesters, and label sites with the feature's accompanying manure management emissions factor may improve the model's performance in regions with more heterogeneous production systems. Feature identification will also improve the accuracy of facility area estimates. Figure 6 shows that some RAIC identified feedlots do not have their full area mapped. As a next step, RAIC will continue refining this approach. This includes providing satellite derived vegetation indices to separate crop areas from the convex hull drawing process and road feature identification to separate convex hull drawings that extend beyond a facility area.

The results reported on the Climate TRACE website results may be adjusted to account for differences in facility utilization, by comparing the estimated aggregate feedlot output (at 100% utilization) of identified livestock facilities to beef processing data and dairy production numbers. Adjustment factors may be incorporated should the estimates overestimate annual production. In areas where livestock spend less time on confined feedlots relative to open pasture, the current approach will fail to capture livestock emissions that originate upstream of feedlots in the supply chain. Countries with predominantly pasture-based systems, tend to still finish beef cattle on concentrated rations in feedlots even if cattle spend less time on those facilities, thus it may still be possible to reverse engineer emissions estimates from grazing operations by integrating the region's total feedlot capacity, the utilization rates of those facilities, the prevailing practices for the number of days cattle are on pasture before a feedlot, and the appropriate emissions factor for grazing in that area. Nonetheless, in places with smallholder cattle production systems that do not utilize feedlots, it will be very difficult to utilize remote sensing approaches like RAIC to assess cattle population and methane emissions. Mass balance or survey approaches may be the most appropriate available methods in those contexts.

The next step with this research will be to expand deployment of the RAIC model internationally. Climate TRACE has engaged Synthetaic to assess the viability of the RAIC tool in dominant beef and dairy production countries, including China, India, Brazil, the European Union, Mexico, Australia, and the remaining unmodeled regions in the USA and Argentina. Climate TRACE will, in parallel, collect ground-truthed, facility level population data to help tune the model to local practices, and test whether other variables may impact cattle density and warrant incorporation into the model. Future work may also integrate practice identification with remote sensing or even top down methane measurements, from providers such as AVIRIS and GHGSat, to help improve emissions modeling at identified facilities, or apply revised emission factors from Wolf et al. (2017) that account for breed and practice changes. The project will also aim to improve temporal resolution by tracking the expansion of facilities over time and referencing historic temperature records to model emissions for previous production years.

Currently, we identified an issue where the 2021 asset emissions are resulting in higher than expected N_2O emissions. We are addressing this issue and will provide updated N_2O emissions that are more reflective of manure management systems employed at beef and dairy feedlots.

6. Acknowledgements

We would like to thank April Chen, Clara Duffy, Leor Fishman, Justas Janonis, Jonas Hansen, and Patrick Song for their engagement in this work through the Harvard Tech for Social Good (socialgood.hcs.harvard.edu).

7. Supplementary material

Table S1 IPCC equations, Emission Factors (EFs), and other variables used to estimate beef and dairy emissions by region.

<table>
<tr><th rowspan="2">IPCC Equations</th><th colspan="2">North America
(California and Texas)</th><th colspan="2">Latin America
(Argentina)</th><th rowspan="2">IPCC Tables</th></tr>
<tr><th>Dairy EFs and other inputs</th><th>Other Cattle (Beef) EFs and other inputs</th><th>Dairy EFs and other inputs</th><th>Other Cattle (Beef) EFs and other inputs</th></tr>
<tr><td>10.19 ENTERIC FERMENTATION EMISSIONS FROM A LIVESTOCK CATEGORY</td><td>128</td><td>53</td><td>72</td><td>56</td><td>10.11</td></tr>
<tr><td>10.22 CH_4 EMISSIONS FROM MANURE MANAGEMENT</td><td>48 - 112</td><td>1 - 2</td><td>1-2</td><td>1</td><td>10.14</td></tr>
<tr><td>10.25 DIRECT N_2O EMISSIONS FROM MANURE MANAGEMENT</td><td>0.005 (Applied liquid/slurry)</td><td>0.02 (Assume Dry lot)</td><td>0.02 (Assume Dry lot)</td><td>0.02 (Assume Dry lot)</td><td>10.21</td></tr>
<tr><td rowspan="2">10.26 N LOSSES DUE TO VOLATILISATION FROM MANURE MANAGEMENT

10.27 INDIRECT N_2O EMISSIONS DUE TO VOLATILISATION OF N FROM MANURE MANAGEMENT</td><td colspan="4">0.010 (EF_4) and 0.0075 (EF_5)</td><td rowspan="4">10.22, 10A-4, 10A-5, 11.3.</td></tr>
<tr><td>$MS_{(T,S)} = 0.15$
$Frac_{GasMS} = 35\%$</td><td>$MS_{(T,S)} = 0.15$
$Frac_{GasMS} = 30\%$</td><td>$MS_{(T,S)} = 0$
$Frac_{GasMS} = 30\%$</td><td>$MS_{(T,S)} = 0$
$Frac_{GasMS} = 20\%$</td></tr>
<tr><td rowspan="2">10.28 N LOSSES DUE TO LEACHING FROM MANURE MANAGEMENT SYSTEMS

10.29 INDIRECT N_2O EMISSIONS DUE TO LEACHING FROM MANURE MANAGEMENT</td><td colspan="4">0.010 (EF_4) and 0.0075 (EF_5)</td></tr>
<tr><td>$MS_{(T,S)} = 0.15$
$Frac_{LeachMS} = 3\%$</td><td>$MS_{(T,S)} = 0.15$
$Frac_{LeachMS} = 3\%$</td><td>$MS_{(T,S)} = 0$
$Frac_{LeachMS} = 3\%$</td><td>$MS_{(T,S)} = 0$
$Frac_{LeachMS} = 3\%$</td></tr>
</table>

8. References

1. Costa Jr, C., Li, C., Cerri, C. E., & Cerri, C. C. (2014). Measuring and modeling nitrous oxide and methane emissions from beef cattle feedlot manure management: First assessments under Brazilian condition. *Journal of Environmental Science and Health, Part B*, *49*(9), 696-711.
2. Euken, R., Doran, B., Clark, C., Shouse, S., Ellis, S., Loy, D. and Schulz, L., 2015. Beef feedlot systems manual PM 1867. *Iowa State University, Ames, IA*.
3. Krekelberg, E. 2020. *Space requirements for dairy cows.* https://extension.umn.edu/dairy-milking-cows/space-requirements-dairy-cows, accessed 30-Aug-2022.
4. *Major cuts of greenhouse gas emissions from livestock within reach*. FAO. (2013, September 26). Retrieved September 19, 2022, from https://www.fao.org/news/story/en/item/197608/icode
5. McFarland, D. and Tyson, J. 2016. *Designing and Building Dairy Cattle Freestalls.* https://extension.psu.edu/designing-and-building-dairy-cattle-freestalls, accessed 20-Aug-2022.
6. Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T.P., Vanwambeke, S.O., Wint, G.R. and Robinson, T.P., 2018. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Scientific data*, *5*(1), pp.1-11.
7. Harper, L.A., Flesch, T.K., Powell, J.M., Coblentz, W.K., Jokela, W.E. and Martin, N.P., 2009. Ammonia emissions from dairy production in Wisconsin. *Journal of Dairy Science*, *92*(5), pp.2326-2337.
8. Harter, T., J. H. Viers, D. Liptzin, T. S. Rosenstock, V. B. Jensen, A. D. Hollander, A. McNally, A. M. King, G., Kourakos, E. M. Lopez, N. DeLaMora, A. Fryjoff-Hung, K. N. Dzurella, H. Canada, S. Laybourne, C., McKenney, J. Darby, and J. F. Quinn, 2013. Dairies and other sources of nitrate loading to groundwater. Task Report 6. SWRCB Agreement Number 04-184-555. Department of Land, Air, and Water Resources. University of California, Davis. 103 pages. http://groundwater.ucdavis.edu.
9. IPCC (Intergovernmental Panel on Climate Change), 2006a. Chapter 10. Emissions from livestock and manure management. *Guidelines for National Greenhouse Inventories. Vol. 4. Agriculture, Forestry and Other Land Use*.
10. IPCC (Intergovernmental Panel on Climate Change), 2006b. Chapter 11. N2O emissions from managed soils, and CO2 emissions from lime and urea application. *Guidelines for National Greenhouse Inventories. Vol. 4. Agriculture, Forestry and Other Land Use*.
11. Owen, J.J. and Silver, W.L., 2015. Greenhouse gas emissions from dairy manure management: a review of field‐based studies. *Global change biology*, *21*(2), pp.550-565.

12. Petersen, S.O., 2018. Greenhouse gas emissions from liquid dairy manure: Prediction and mitigation. *Journal of dairy science*, *101*(7), pp.6642-6654.
13. Waldrip, H.M., Todd, R.W., Parker, D.B., Cole, N.A., Rotz, C.A. and Casey, K.D., 2016. Nitrous oxide emissions from open‐lot cattle feedyards: a review. *Journal of environmental quality*, *45*(6), pp.1797-1811.
14. Waldrip, H.M., Parker, D.B., Miller, S., Miller, D.N., Casey, K.D., Todd, R.W., Min, B.R., Spiehs, M.J. and Woodbury, B., 2020. Nitrous Oxide from Beef Cattle Manure: Effects of Temperature, Water Addition and Manure Properties on Denitrification and Nitrification. *Atmosphere*, *11*(10), p.1056.
15. Wolf, J., Asrar, G.R. and West, T.O., 2017. Revised methane emissions factors and spatially distributed annual carbon fluxes for global livestock. *Carbon balance and management*, *12*(1), pp.1-24.
16. Zhu, G., Ma, X., Gao, Z., Ma, W., Li, J. and Cai, Z., 2014. Characterizing CH4 and N2O emissions from an intensive dairy operation in summer and fall in China. *Atmospheric Environment*, *83*, pp.245-253.