California Dairy and Livestock Database (CADD)

Version 1.1.0

Technical Document

July 2025

Atmospheric Processes Research Section
Atmospheric Science and Climate Strategies Branch
Research Division

Table of Contents

1 Introduction 2
2 Database Structure 2
2.1 Overview 2
2.2 Table Relationships 3
3 Description of Data Tables 3
3.1 Facility General Information Table 3
3.2 Facility Herd Size Table 4
3.3 Anaerobic Digesters Table 5
4 Facilities Included in CADD 5
5 CADD Development 6
5.1 Facility General Information Table 6
5.2 Facility Herd Size Table 7
5.3 Anaerobic Digesters Table 16
6 Comprehensiveness of CADD for Dairies 17
7 Summary 18
8 Appendix A – Reference Code for Herd Size Data 20
9 Appendix B – Regional Water Quality Control Boards' Orders Impacting Cattle Facilities 22
10 Appendix C – Identifying Potentially Operational Facilities 26
11 Appendix D – Decision Tree to Address Missing Herd Size Data 27

1 Introduction

California's dairy and livestock sector contributes to 54% of the statewide methane (CH_4) emissions.[1] To reduce these emissions, California's Legislature passed Senate Bill 1383 (SB 1383, Lara, Chapter 395) in 2016, which requires the California Air Resources Board (CARB) to approve and implement strategies to reduce CH_4 emissions from the dairy and livestock sector by 40% below 2013 levels by 2030. To enhance the understanding of emissions trends in this sector, CARB developed the California Dairy and Livestock Database (CADD) using data from Federal, State, and Local agencies. The database provides facility-level cattle herd size and anaerobic digester information for California dairies from 2012 to 2022. CADD is based on publicly accessible documents and improves CARB's ability to track trends in the state's dairy and livestock sector. At this time, this database is the most comprehensive collection of dairy and livestock data for the state, pulling from several publicly available data sources.

CARB released the first version of CADD on August 2, 2024, and presented it to the public during the Dairy Sector Workshop[2] on August 22, 2024, seeking public feedback. Following the comment period, CARB incorporated stakeholders' input and updated CADD to version 1.1.0 by significantly enhancing its transparency. This technical document outlines CADD's data content, the types of facilities included, and its development process. It also compares CADD's data coverage to other data sources. CADD is available on the CARB website at *https://ww2.arb.ca.gov/resources/documents/california-dairy-livestock-database-cadd*. For additional information about this database, please contact *cadd@arb.ca.gov*.

2 Database Structure

The database is composed of three interconnected tables: the **Facility General Information table,** the **Facility Herd Size table,** and the **Anaerobic Digesters table**.

2.1 Overview

1. **Facility General Information Table**: Stores information about each facility in CADD, including unique identifiers, names, location information, and the associated Regional Water Quality Control Boards.
2. **Facility Herd Size Table**: Contains herd size data of seven cattle categories (described in Section 3.2) for all facilities in the Facility General Information table for every year from 2012 to 2022. Herd size data for some non-dairy cattle facilities in CADD spans from 2019 to 2022 (refer to Section 4 for additional information).
3. **Anaerobic Digesters Table**: Contains facilities with anaerobic digesters or facilities that have an anaerobic digester in some stage of planning, financing, or construction, the operational year (or the expected operational year) of the digesters, and the year they shut down (if applicable).

[1] *California AB 32 GHG Emissions Inventory 2000-2022 (2024 Edition) | California Air Resources Board*
[2] *Dairy Sector Workshop | California Air Resources Board*

2.2 Table Relationships

Figure 1 shows the relationships between different tables. `CADDID`, a unique identifier associated with each facility in CADD, is used across all tables to connect information. While each `CADDID` appears only once in the Facility General Information table, it can appear multiple times in the other two tables.

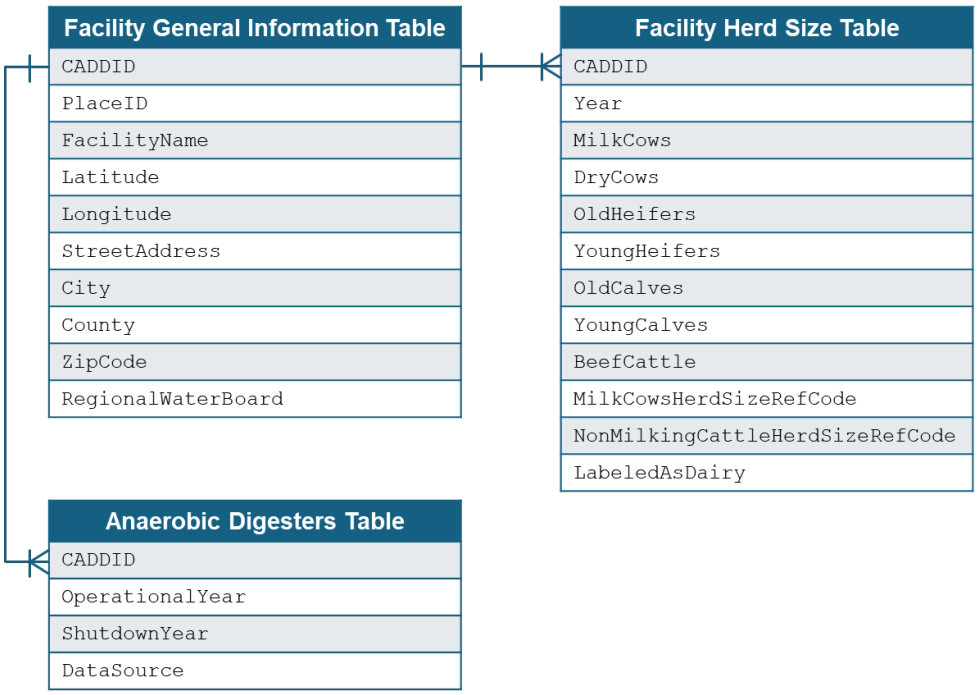

Figure 1. Relationships between different tables in CADD.

3 Description of Data Tables

3.1 Facility General Information Table

The Facility General Information table stores information about each facility in CADD, including unique identifiers, names, location information, and the associated Regional Water Quality Control Boards. For information about what facilities are included in CADD, refer to Section 4.

The columns in this table are listed below:

- `CADDID`: Whole number (integer) – a unique identifier associated with each facility in CADD,
- `PlaceID`: Whole number (integer) – Place ID of the corresponding facility in the State Water Board system, if applicable,
- `FacilityName`: Text – Name of each facility; since facilities may change their names frequently, `FacilityName` may not reflect the current name,
- `Latitude`: Decimal (double) – Geographic coordinate specifying the north-south position of each facility on the Earth's surface, expressed in decimal degrees,
- `Longitude`: Decimal (double) – Geographic coordinate specifying the east-west position of each facility on the Earth's surface, expressed in decimal degrees,

- `StreetAddress`: Text – The detailed location of each facility, including building number and street name,
- `City`: Text – The city associated with each facility,
- `County`: Text – The county associated with each facility,
- `ZipCode`: Whole number (integer) – A five-digit zip code associated with each facility,
- `RegionalWaterBoard`: Text – The Regional Water Quality Control Board associated with each facility.

3.2 Facility Herd Size Table

The Facility Herd Size table contains average herd size data of seven cattle categories for all facilities in the Facility General Information table, for every year from 2012 to 2022. Herd size data for some non-dairy cattle facilities in CADD spans from 2019 to 2022 (see Section 4 for more information). The seven cattle categories include: Milk Cows, Dry Cows, Old Heifers (Heifers 15–24 months old), Young Heifers (Heifers 7–14 months old), Old Calves (Calves 4–6 months old), Young Calves (Calves 0–3 months old), and Beef Cattle.

The columns in this table are listed below:

- `CADDID`: Whole number (integer) – a reference to the facility; linked to the `CADDID` in the Facility General Information table,
- `Year`: Whole number (integer) – the calendar year the herd size is associated with,
- `MilkCows`: Whole number (integer) – average milk cow herd size of `CADDID` during the `Year`,
- `DryCows`: Whole number (integer) – average dry cow herd size of `CADDID` during the `Year`,
- `OldHeifers`: Whole number (integer) – average old heifer herd size of `CADDID` during the `Year`,
- `YoungHeifers`: Whole number (integer) – average young heifer herd size of `CADDID` during the `Year`,
- `OldCalves`: Whole number (integer) – average old calf herd size of `CADDID` during the `Year`,
- `YoungCalves`: Whole number (integer) – average young calf herd size of `CADDID` during the `Year`,
- `BeefCattle`: Whole number (integer) – average beef cattle herd size of `CADDID` during the `Year`,
- `MilkCowsHerdSizeRefCode`: Whole number (integer), may be followed by a letter – indication of how milk cow herd size information is determined:
    - One (1): herd size information is based on an annual report submitted to the Regional Water Quality Control Boards,
    - Begins with two (2): no annual report found for `CADDID` during the `Year` and the facility is deemed not to be an active milking operation. The letter indicates what data source was used for this determination. Refer to Section 5.2.4 and Appendix A for additional information,
    - Begins with three (3): no annual report found for `CADDID` during the `Year` but the facility is deemed to be an active milking operation. The letter indicates what data source was used to estimate milk cow herd size data. Refer to Section 5.2.4 and Appendix A for additional information.
- `NonMilkingCattleHerdSizeRefCode`: Whole number (integer), may be followed by a letter – indication of how non-milking cattle herd size information is determined:

    - One (1): herd size information is based on an annual report submitted to the Regional Water Quality Control Boards,
    - Begins with two (2): no annual report found for `CADDID` during the `Year` and the facility is deemed to not house non-milking cattle. The letter indicates what data source was used for this determination. Refer to Section 5.2.4 and Appendix A for additional information,
    - Begins with three (3): no annual report found for `CADDID` during the `Year` but the facility is deemed to house non-milking cattle. The letter indicates what data source was used to estimate non-milking cattle herd size data. Refer to Section 5.2.4 and Appendix A for additional information.
- `LabeledAsDairy`: Whole number (integer) – an indication of whether or not `CADDID` housed milk cows during the `Year`:
    - Zero (0): Labeled as not a dairy since `MilkCows` is equal to zero for `CADDID` during the `Year`,
    - One (1): Labeled as a dairy since `MilkCows` is greater than zero for `CADDID` during the `Year`.

3.3 Anaerobic Digesters Table

Anaerobic Digesters table contains facilities with anaerobic digesters or facilities that have an anaerobic digester in some stage of planning, financing, or construction, the year when they became operational, and the year they shut down (if applicable).

The columns in this table are listed below:

- `CADDID`: Whole number (integer) – a reference to the facility; linked to the `CADDID` in the Facility General Information table,
- `OperationalYear`: Whole number (integer) – the calendar year in which the anaerobic digester became operational or is expected to become operational. For California Department of Food and Agriculture (CDFA) Dairy Digester Research and Development Program (DDRDP) projects, `NaN` means “verification in process”; for non-DDRDP projects, `NaN` means the digester is under construction,
- `ShutdownYear`: Whole number (integer) – the calendar year in which the anaerobic digester shut down, if applicable. `NaN` means the digester has not shut down (either is not constructed yet or is operational),
- `DataSource`: Text – the data source for each record: DDRDP or AgSTAR (described in Section 5.3).

4 Facilities Included in CADD

Facilities in CADD include cattle facilities that are subject to reporting to the Regional Water Quality Control Boards (refer to Table S1) as well as those that a) received funding from Alternative Manure Management Program (AMMP) or DDRDP (as of March 2023), b) have been issued air permits as of 2019 by the San Joaquin Valley Air Pollution Control District (SJVAPCD), or c) housed cattle any time between 2012 and 2022 based on Google Earth imagery; individual facilities were investigated using their Google Earth images if they appeared in Vista-CA or if the California Integrated Water Quality System (CIWQS) database suggested that the facility may have been operational between 2012 and 2022 (see Appendix C for more details). If Google Earth images suggested that animals existed on the

facility at any time between 2012 and 2022, the facility was added to CADD. Information on data sources used to compile the list of facilities in CADD can be found in Section 5.

While CADD focuses on dairies, it also includes information about non-dairy cattle facilities. Some non-dairy cattle facilities began submitting annual reports in 2019, as required by the Central Valley Regional Water Quality Control Board (R5) Confined Bovine Feeding Operations General Order (R5-2017-0058). If a facility did not submit an annual report before 2019 and a) its "Order #" in the CIWQS database is "R5-2017-0058", or b) appeared in the Vista-CA Feed Lot layer, its herd size data spans from 2019 to 2022. Otherwise, it is 2012 to 2022. While CADD contains valuable facility-level herd size information for the regulated facilities, since some non-dairy cattle facilities such as grazing operations are exempt from reporting under the Confined Bovine Feeding Operations General Order, CADD's current version does not capture all non-dairy cattle facilities in the state.

5 CADD Development

This section describes the steps involved in developing CADD.

5.1 Facility General Information Table

The sources of data used to develop this table are listed below:

- **California Integrated Water Quality System (CIWQS) from State Water Board**:[3] CIWQS, managed by the California State Water Board, is a computer system that allows the State and Regional Water Quality Control Boards to track information about places of environmental interest, manage orders, track inspections, and manage violations and enforcement activities. CIWQS contains general information about facilities, such as unique identifiers (Place ID), facility names, and location information. CIWQS was the primary dataset used to construct the Facility General Information table. CIWQS also reports the Confined Animal Facility Operations (CAFO) Population, which is primarily used for billing purposes. While, in general, values in CIWQS CAFO Population adequately represent herd sizes, they do not match herd sizes reported in annual reports (results not shown). Therefore, although it may be appropriate to use values in the CIWQS CAFO Populations for billing purposes, its inclusion in CADD would have led to inaccuracies.
- **California Climate Investment (CCI) Implemented Projects Database from CARB**:[4] provides detailed information on facilities that have received funding from several State programs, such as AMMP and DDRDP. The database includes facility names and location information. The location information of the DDRDP and AMMP awardees in CADD was manually verified by comparing it with the location information reported in the CCI-implemented projects database for DDRDP and AMMP.
- **Vista-CA**:[5] consists of detailed spatial maps for facilities and infrastructure in California that are known (or expected) sources of CH_4 emissions. This dataset was created using publicly available data sources from Local, State, and Federal agencies for the years 2005–2019. Three Google Earth layers (Dairies, Feed Lots, and Digesters) from Vista-CA were used in CADD; the

[3] *Surface Water - Water Quality Regulated Facility Information - Confined Animal Facilities Regulated by the Water Boards - California Open Data* (last accessed November 21, 2023).
[4] *CCIRTS GGRF Application (arcgis.com)* (last accessed October 23, 2023).
[5] *Sources of Methane Emissions (Vista-CA), State of California, USA, https://doi.org/10.3334/ORNLDAAC/1726* (last accessed December 28, 2023).

closest CADD facility to each Vista-CA facility was identified and the Vista-CA facility was added to CADD if it was not already included.

While initial location information for CADD was sourced from CIWQS, additional measures were implemented to verify this information. The GeoCode[6] add-on in Google Sheets was used to convert each CADD facility's CIWQS address to latitude and longitude. The GeoCode's latitude and longitude were compared to the CADD facility's CIWQS latitude and longitude. If the distance between GeoCode's values and CIWQS's exceeded 800 m, the CIWQS's location information was manually investigated to verify and update the records. The 800-meter threshold was selected because a manual investigation of CIWQS's values for facilities with distances between 500 m and 800 m confirmed that the CIWQS's values were accurate.

CADD's location information was further refined by integrating data from the CCI Implemented Projects Detailed Database. The CCI facility-level location information was compared to CADD's, and if CADD's location information did not match CCI's, it was manually investigated to verify and update records.

Vista-CA's location information was also utilized to improve CADD's location information. The closest CADD facility to each Vista-CA facility was identified and the location information was manually investigated to verify and update records if the distance between the Vista-CA facility and its matching CADD facility exceeded 500 meters or if one CADD facility was linked to more than one Vista-CA facility.

5.2 Facility Herd Size Table

This section describes the steps involved in developing the Facility Herd Size table. A visual representation of these steps is outlined in Figure 2.

Figure 2. Visual representation of the steps involved in developing the Facility Herd Size table.

5.2.1 Public Data Retrieval

In the first step of Facility Herd Size table development, CARB staff retrieved relevant public data sources from several Federal, State, and Local agencies focusing on data sources since 2012. The total number of public documents used to develop this version of CADD exceeded 12,000. This effort included multiple visits by CARB staff to the Regional Water Quality Control Board offices to manually review and digitize thousands of paper records.

The sources of data used to develop this table are listed below:

- **Annual reports from Regional Water Quality Control Boards:** In California, all owners/operators of cattle facilities, with some exemptions, are required to submit annual reports to their regulating Regional Water Quality Control Boards. These reports include average herd size information for multiple animal types, such as milk cows, dry cows, heifers,

[6] https://workspace.google.com/marketplace/app/geocode_by_awesome_table/904124517349

calves, and beef cattle. These reports served as the primary data source for CADD herd size data due to their comprehensiveness (statewide), frequency (annual), historical record, and detailed cattle categories. These reports were retrieved by CARB staff through communication with the Regional Water Quality Control Boards' staff. A comprehensive list of Regional Water Quality Control Boards' orders governing dairy facilities and non-dairy cattle facilities, along with their exemptions, is presented in Table S1. In addition, Table S2 provides an overview of the reporting years for each Regional Water Quality Control Board.[7]

- **Inspection reports**:[8] conducted by Regional Water Quality Control Boards' staff and uploaded to the CIWQS website; these reports typically provide the facility's herd size data at the time of inspection, serving as an alternative source of information in case of a missing annual report.
- **Air permits**:[9] issued by the SJVAPCD; these permits specify the maximum permitted herd size for each SJVAPCD-permitted facility, serving as an alternative source of information in case of a missing annual report.
- **Regulation information from CIWQS**:[10] provides information about the operational status of a facility in a given year in case of a missing annual report.[11]
- **Google Earth images**: used to evaluate the operational status of a facility in a given year in case of a missing annual report. Since Google Earth provides historical images from different years, the closest available images that best represented the facility's operational status during the year of interest were used.
- **Waste management plans, nutrient management plans, and notices of intent**: These documents are generally submitted to the Regional Water Quality Control Boards once by the facility owners/operators to enroll in the Dairy or Confined Bovine Feeding Operations General Orders. Typically containing the facility's herd size data at the time of document submission, they serve as alternative sources of information in case of a missing annual report. These documents were retrieved by CARB staff through communication with the Regional Water Quality Control Boards' staff.

5.2.2 Enter Herd Size Data into MS Access

Annual reports from different Regional Water Quality Control Boards follow different reporting formats. In addition, while some Regional Water Quality Control Boards provided actual annual reports in a PDF format, others supplied spreadsheets listing herd size information for their regulated facilities (Table S3). In cases where spreadsheets were provided, the data was directly transferred to an MS Access database. For instances without available spreadsheets (most of the data), herd size data were extracted directly from annual report PDFs. The extraction process involved using a Python module and manual entry into the MS Access database.

5.2.2.1 Python PDF Reader

Leveraging Python as a PDF reader, specialized modules using the Optical Character Recognition (OCR) technology were developed to extract herd size information from annual reports, particularly those submitted by facilities falling under the jurisdiction of the Central Valley Regional Water Quality

[7] No cattle facility is regulated/identified in the Los Angeles Regional Water Quality Control Board (Region 4).
[8] *California Integrated Water Quality System* (last accessed February 16, 2024).
[9] *Permits Public Information Portal (valleyair.org)* (last accessed July 3, 2019).
[10] *Surface Water - Water Quality Regulated Facility Information - Confined Animal Facilities Regulated by the Water Boards - California Open Data* (last accessed November 21, 2023).
[11] Includes information on the "order number" a facility is subject to, the effective date of the regulation measure, the termination date of the order for that facility (if applicable), etc.

Control Board. This automated approach expedited the data extraction process, presenting a substantial time-saving alternative to manual data extraction. To ensure all raw data from annual reports were stored in a single location, herd size data extracted by the Python modules was transferred to the MS Access database tables.

5.2.2.2 Manual Data Extraction

For annual reports from which herd size information couldn't be extracted using the Python modules (mostly due to poor resolution of the scans), a manual extraction process was employed. MS Access forms, resembling the annual report formats, were devised for data entry, and the entered data was stored in corresponding MS Access database tables. These forms minimized data entry errors (Figure 3). While annual reports typically contain additional information such as generated waste, crop, and manure land application, our focus remained solely on extracting herd size information.

a)

HERD INFORMATION

	Milk Cows	Dry Cows	Bred Heifers (15-24 mo.)	Heifers (7-14 mo. to breeding)	Calves (4-6 mo.)	Calves (0-3 mo.)
Number open confinement	0	150	0	0	0	0
Number under roof	1,780	0	0	0	0	0
Maximum number	1,800	160	0	0	0	0
Average number	1,790	155	0	0	0	0
Avg live weight (lbs)	1,400	1,450	0	0		

Predominant milk cow breed: Holstein

Average milk production: 62 *pounds per cow per day*

b)

HERD INFORMATION

Table Populated? 1	Milk Cows	Dry Cows	Bred Heifers (15-24 mo.)	Heifers (7-14 mo. to breeding)	Calves (4-6 mo.)	Calves (0-3 mo.)
Number open confinement	0	150	0	0	0	0
Number under roof	1,780	0	0	0	0	0
Maximum number						
Average number						
Avg live weight (lbs)	1,400	1,450	0	0		

Predominant milk cow breed: 1

Average milk production: 62 pounds per cow per day

Record: 1 of 1 No Filter Search

Figure 3. An example of an MS Access form resembling the annual report herd size information table. a) a snapshot of a sample annual report and b) the MS Access form used for manual data entry.

5.2.3 Break Down Coarse Herd Size Information into Detailed Cattle Categories

Herd size data in the MS Access database was copied to a spreadsheet for further processing. At this stage, CADD included over 16,000 records, each containing herd sizes of the reported cattle categories for any year between 2012 and 2022 by relying on the provided annual reports. However, the cattle categories listed in annual reports are not consistent across different Regional Water Quality Control Boards. For example, while some regions may focus solely on dairies' mature cows (milk cows + dry cows) and non-mature cows, others may focus on more detailed cattle categories such as milk cows, dry cows, old heifers, young heifers, old calves, and young calves (Table S3). It is important to note that, on average, larger dairies reported detailed herd size information. For example, while

approximately 12% of the records reported coarse herd size information for at least one cattle category, their contribution to the statewide milk cow population is much smaller, at about 3.5%.

To ensure that all Facility Herd Size table data records have consistent cattle categories, the reported coarse herd size information was broken down into detailed cattle categories using average ratios of herd sizes of different cattle categories, as shown in Table 1. These average ratios were calculated using the annual reports that reported the most detailed cattle categories. The annual report formats that provided the most detailed herd size information for dairies included Merced Tool, Innovative Ag (Dairy General Order version), and Provost & Pritchard (P&P). These formats reported herd sizes for milk cows, dry cows, old heifers (15–24 months old), young heifers (7–14 months old), old calves (4–6 months old), and young calves (0–3 months old). To derive average herd size ratios of different cattle categories for dairies, staff used data that was extracted from these formats for facilities that reported milk cow herd sizes of greater than 2.[12] The average ratios of herd sizes for different cattle categories for dairies are presented in Table 1.

As an example of how these ratios were used, consider a dairy that reports 100 mature cows and 100 non-mature cows. Applying the ratios from Table 1, the estimated herd size for this dairy would be 88 milk cows, 12 dry cows, 42 old heifers, 30 young heifers, 17 old calves, and 11 young calves. This methodology ensures a standardized representation of detailed herd sizes across all facilities in CADD.

Table 1. Calculated average ratios of herd sizes of different cattle categories from annual reports that reported detailed cattle categories (dairies).

Numerator / Denominator	Milk Cows	Dry Cows	Old Heifers	Young Heifers	Old Calves	Young Calves
Mature Cows	0.88	0.12				
Heifers			0.58	0.42		
Calves					0.68	0.32
Non-Mature Cows			0.42	0.30	0.17	0.11
Non-Milking Cows*		0.25	0.30	0.23	0.13	0.08
Total Cows	0.57	0.08	0.14	0.11	0.06	0.04

* Totals may not add up to 1 due to rounding.

The annual report formats that provided the most detailed herd size information for non-dairy cattle facilities included Merced Tool, Innovative Ag (Dairy General Order version and Confined Bovine

[12] The threshold of 2 for milk cows herd size was chosen as some facility operators/owners noted that Merced Tool does not generate an annual report if the entered milk cow herd size is below this threshold.

Feeding Operations General Order version), and Bovine Format (Confined Bovine Feeding Operations Only). These formats reported herd sizes for beef cattle, dry cows, old heifers, young heifers, old calves, and young calves.[13] To derive average herd size ratios of different cattle categories for non-dairy cattle facilities, the data extracted from these formats was filtered for facilities that either submitted annual reports in formats specific to non-dairy cattle facilities (Innovative Ag (Confined Bovine Feeding Operations General Order version) and Confined Bovine Feeding Operations Format) or their reported milk cow herd size was less than or equal to 2 and their total herd size was at least 5.[14] The average ratios of herd sizes for different cattle categories for non-dairy cattle facilities are presented in Table 2.

Table 2. Calculated average ratios of herd sizes of different cattle categories from annual reports that reported detailed cattle categories (non-dairy cattle facilities).

Numerator / Denominator	Beef Cattle	Dry Cows	Old Heifers	Young Heifers	Old Calves	Young Calves
Heifers			0.49	0.51		
Calves					0.82	0.18
Total Cows	0.18	0.02	0.23	0.25	0.23	0.09

5.2.4 Addressing Missing Herd Size Data

At this stage, the Facility Herd Size table included the herd sizes for seven cattle categories for each facility for any year between 2012 and 2022 (herd size data for some non-dairy cattle facilities in CADD spans from 2019 to 2022) by relying on the herd size data provided in the annual reports. If an annual report was not submitted for a year, the table had a missing data point. Reasons for missing data points included the facility not being operational for that year, the facility being exempt from a Regional Water Quality Control Board order, the facility being non-compliant with the Regional Water Quality Control Boards' orders leading to the non-submission of the annual report, or the Regional Water Quality Control Board not collecting data for that specific year. It's also important to note that some facilities may not have been operational since 2012, and therefore, no documentation on their herd size may be available.

In cases where an entire region had missing herd size data for specific years, herd size values for adjacent years were used to address missing data points:

- For Region 1, where non-mature cows were not reported for 2020, herd sizes were interpolated using data from 2019 and 2021.
- For Region 2, where herd size data was not available until 2016, values from 2016 were assumed for prior years.

[13] Merced Tool and Innovative Ag (Dairy General Order version) formats do not report beef cattle but report milk cows.

[14] See footnote 12.

- For Region 7, where herd size data for 2012, 2013, and 2017 was missing, values from 2014 were used for 2012 and 2013, and interpolation between 2016 and 2018 was applied for 2017.

The remaining missing herd sizes were addressed by developing and following a comprehensive decision tree (Figure S1). This approach allowed for the systematic handling of complex facility-level scenarios. Given that some facilities may cease dairy operations but continue to house non-milking cattle, separate methods were used to address herd sizes for milk cows and non-milking cattle.

In response to public comments requesting greater data transparency in the first version of CADD, staff added two columns to the Facility Herd Size table in CADD v1.1.0. These columns specify the data source for milk cow and non-milking cattle herd size data for every record in CADD. The process to address missing herd size data is described below, and the codes used in these two columns are described in Appendix A. It is important to note that these improvements did not change the herd size values from CADD v1.0.0.

5.2.4.1 Milk Cow Herd Size Data

To address missing milk cow herd size data, staff first determined whether the dairy in question was an active milking operation during the specified year. To make this determination, evidence (in the order of priority) from the following sources was reviewed:

- Water Board records including inspection reports and notes in CIWQS regulation information,
- Additional Water Board records including notes in annual reports for other years, and notice of intent,
- Google Earth images, or
- Maximum milk cow herd sizes listed in air permits.

If any of this evidence indicated that the dairy was not an active milking operation, its milk cow herd size was set to zero. Otherwise, the milk cow herd size was estimated using herd size data from one of the following alternative sources (in the order of priority):

1. Annual reports for other years with the same owner/operator,
2. Inspection reports for the same year or other years with the same owner/operator,
3. Air permits for the same year or other years with the same owner/operator,
4. Waste management plans for the same year or other years with the same owner/operator,
5. Nutrient management plans for the same year or other years with the same owner/operator,
6. Notice of intent for the same year or other years with the same owner/operator.

In rare cases, a dairy determined to be operational may have its milk cow herd size set to zero due to the lack of herd size data in all available data sources.

5.2.4.2 Non-Milking Cattle Herd Size Data

To address missing milk cow herd size data, staff first determined whether the facility in question housed non-milking cattle during the specified year. To make this determination, evidence (in the order of priority) from the following sources was reviewed:

- Water Board records including inspection reports and notes in CIWQS regulation information,
- Additional Water Board records including notes in annual reports for other years, and notice of intent, or
- Google Earth images

If the evidence indicated that the facility did not house non-milking cattle, its non-milking cattle herd size was set to zero. Otherwise, the non-milking cattle herd size was estimated using herd size data from one of the following alternative sources (in the order of priority):

1. Annual reports for other years with the same owner/operator,
2. Inspection reports for the same year or other years with the same owner/operator,
3. Air permits for the same year or other years with the same owner/operator,
4. Waste management plans for the same year or other years with the same owner/operator,
5. Nutrient management plans for the same year or other years with the same owner/operator,
6. Notice of intent for the same year or other years with the same owner/operator.

In rare cases, a facility determined to house non-milking cattle may have its non-milking cattle herd size set to zero due to the lack of herd size data in all available data sources.

5.2.4.3 Using Air Permits Data to Address Missing Herd Size Data

In the case of using air permit maximum herd size data to address a missing herd size, an analysis of dairies with available 2019 annual reports and air permits as of 2019 informed the estimation of average herd size for mature cows and support stock from the air permit's maximum herd size information. As indicated in Figure 4, on average, mature cow herd sizes reported in 2019 annual reports are 75.0% of the maximum mature cow herd sizes in the air permits for 2019. Figure 5 indicates that, on average, support stock herd sizes reported in the 2019 annual report are 62.3% of the maximum support stock herd sizes in the air permits for 2019. Support stock herd size was assumed to be the total of heifers and calves in a facility.

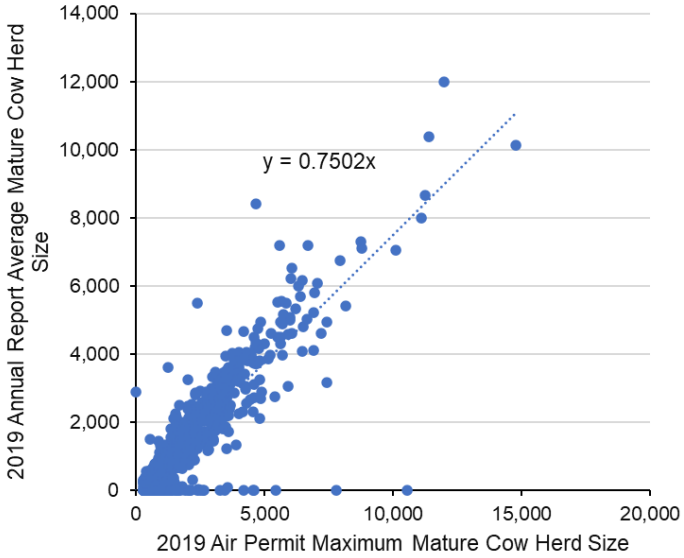

Figure 4. Relationship between annual report average mature cow herd size and air permit maximum herd size.

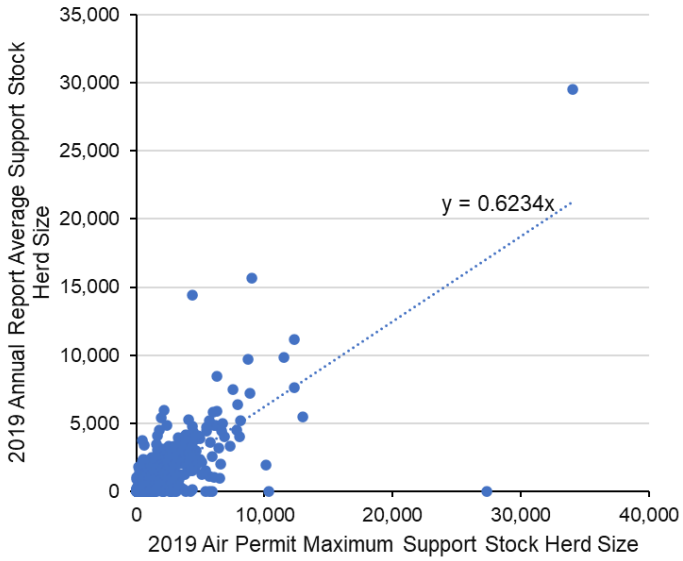

Figure 5. Relationship between annual report average support stock herd size and air permit maximum herd size.

5.2.4.4 Statistics of Missing Herd Size Data in CADD Records

Out of 21,305 records in CADD, 16,189 (76% of the total) directly corresponded to submitted annual reports, while the remaining 5,116 (24% of the total) were initially missing and had to be addressed. The breakdown of these missing records is as follows:

Milk Cows: Of the 5,116 missing records, 3,865 were deemed to be non-operational dairies, resulting in zero milk cows. The remaining 1,251 records were deemed to be operational dairies. Of these, 984 records were resolved using alternative data sources:

- 772 were addressed with annual reports from other years,
- 132 were addressed with inspection reports,
- 22 were addressed with air permits, and
- 58 were addressed using waste management plans, nutrient management plans, and notices of intent.

The remaining 267 records were assigned zero for milk cows since no herd size information was available in any of the data sources; Google Earth images confirmed that all of these records corresponded to relatively small facilities. These statistics are shown in Figure 6.

Non-Milking Cattle: Of the 5,116 records, 2,725 were deemed to not house any cattle, resulting in zero non-milking cattle. The remaining 2,391 records were deemed to house non-milking cattle. Of these, 1,578 records were resolved using alternative data sources:

- 1,213 were addressed with annual reports from other years,
- 255 were addressed with inspection reports,
- 33 were addressed with air permits, and
- 77 records were addressed using waste management plans, nutrient management plans, and notices of intent.

The remaining 813 records were assigned zero for non-milking cattle since no herd size information was available in any of the data sources. These statistics are shown in Figure 6.

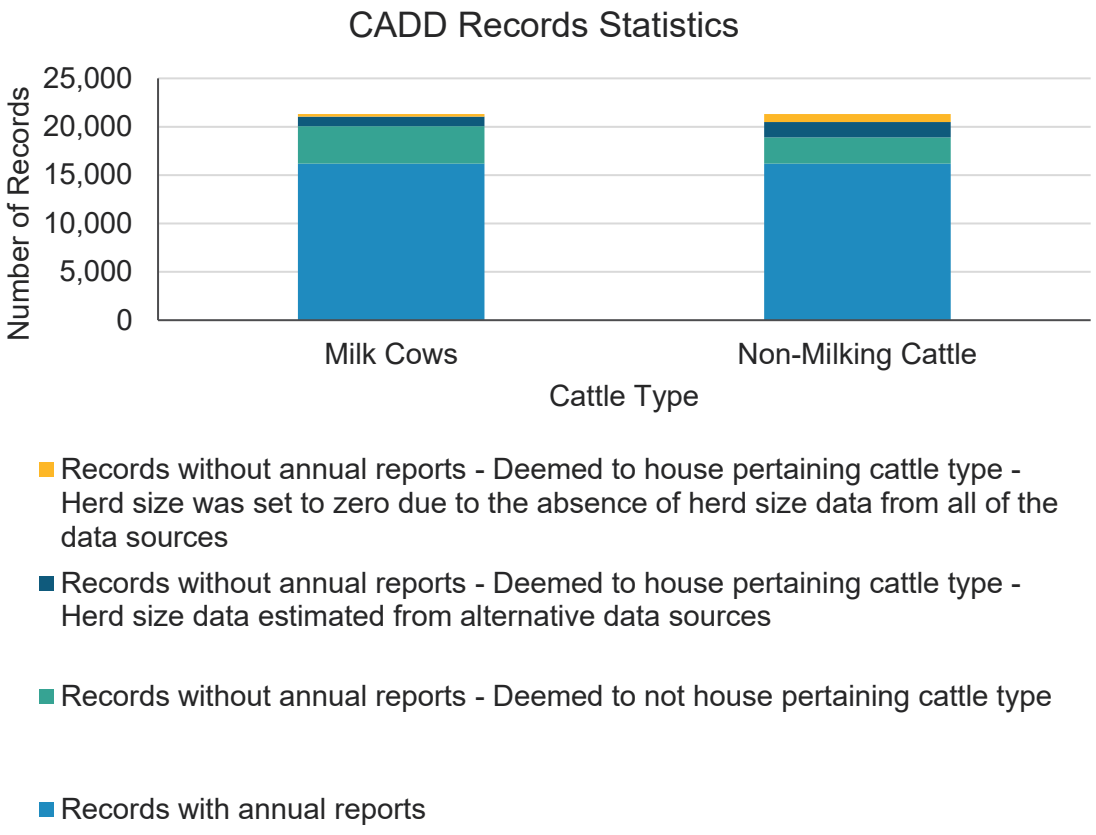

Figure 6. CADD records statistics as they relate to milk cows and non-milking cattle.

5.2.5 Quality Assurance on Herd Sizes

At this stage, the Facility Herd Size table contained herd sizes of detailed animal types at the facility level for all CADD facilities spanning 2012–2022 (herd size data for some non-dairy cattle facilities in CADD spans from 2019 to 2022). This section describes additional measures taken to ensure the accuracy of CADD, focusing on herd sizes.

In case of significant herd size changes of mature cows for each CADD facility, the entered data were manually double-checked and data entry errors were corrected. A significant herd size change of mature cows occurred if the mature cow (milk cow + dry cow) herd size of a facility: 1) turned positive after being zero for at least one year, or 2) was considered an outlier based on the following criteria:

$$MatC_i > \max(\overline{MatC} + 2 \times SD_{MatC}, 1.2 \times \overline{MatC}) \; or \; MatC_i < \min(\overline{MatC} - 2 \times SD_{MatC}, 0.8 \times \overline{MatC})$$

where $MatC_i$ is the facility's mature cow herd size for the year of interest, $\overline{MatC}$ is the facility's average mature cow herd size, excluding the year of interest, and SD_{MatC} is the facility's standard deviation of mature cow herd size, excluding the year of interest.

The same procedure was followed to address potential data entry errors for facilities with significant herd size changes in beef cattle.

Finally, a corrective procedure was applied to ensure an accurate representation of herd sizes and avoid overestimation of herd sizes and the number of operational dairies in erroneous reporting cases. Certain facility operators/owners highlighted in their annual report notes that a commonly used

reporting format utilized by the Central Valley Regional Water Quality Control Board doesn't generate an annual report if the entered milk cow herd size is below 2. In such cases, operators/owners might have entered the value of 1 or 2, even if no milk cows were present. To address potential accounting errors due to this issue, the following procedure was implemented:

1. If reported Milk Cows herd size = 1 or 2, assume Milk Cows = 0.
2. If Dry Cows = 1 or 2 and Milk Cows = 1 or 2, assume Dry Cows = 0.
3. For other cattle categories, if a category's herd size = 1 or 2, and Milk Cows = 1 or 2, and Dry Cows = 1 or 2, assume the category's herd size = 0.

5.3 Anaerobic Digesters Table

The sources of data used to develop this table are listed below:

- **Project-level data from DDRDP**:[15] published quarterly by CDFA, lists information about each DDRDP-funded project, including the award year, facility name, location information, start date of digester construction, and completion date. `OperationalYear` column of CADD for DDRDP-funded projects is based on the completion date in these documents.
- **AgSTAR Livestock Anaerobic Digester database from the United States Environmental Protection Agency (US EPA)**:[16] contains information about the status of dairy digester projects within the United States, including project names, cities, project statuses (operational, under construction, or shutdown), the year of starting operation (for operational and shutdown projects), and the shutdown year, if applicable.

Staff compiled the list of anaerobic digesters using these two data sources, excluding digesters that shut down prior to 2012. In addition, the AgSTAR database was filtered to include only projects located in California.

For DDRDP projects, the completion date was used to determine the `OperationalYear`. Projects with a status of "verification in progress" were not considered operational, and their `OperationalYear` was set to `NaN`. For digesters appearing in AgSTAR, the Year Operational field was used to determine the `OperationalYear`, unless the status was listed as "construction," in which case the `OperationalYear` was set to `NaN`.

The Year Shutdown field in AgSTAR was used to populate the `ShutdownYear` column in the Anaerobic Digesters table. In cases where the same project appeared in both datasets and the reported operational years differed, the value from DDRDP was used.

CADD was finalized at the end of this step. The following section evaluates its comprehensiveness for dairies against other data sources.

[15] *CDFA - OEFI - Dairy Digester (ca.gov)* (published on March 24, 2023 and accessed on April 13, 2023).

[16] *Livestock Anaerobic Digester Database | US EPA* (last accessed on January 10, 2024; based on data available through May 2022).

6 Comprehensiveness of CADD for Dairies

This section compares the number of dairies included in CADD to those permitted by CDFA and those included in the United States Department of Agriculture (USDA) Census of Agriculture (Ag Census) to assess the comprehensiveness of CADD in terms of its coverage of dairy farms.

All dairy farms that produce milk for distribution must hold a permit issued either directly by CDFA or by a county CDFA-approved Milk Inspection Service. A "dairy farm" is defined in the California Food and Agricultural Code in Section 32505 as "any place or premises upon which milk is produced for sale or other distribution and where more than two cows or water buffalo, or six goats, sheep, or other hooved mammals, are in lactation."[17] CARB staff obtained the number of permitted cattle dairy farms in California from CDFA covering the period from 2012 to 2022 (last updated on January 22, 2024). Figure 7 shows that the number of dairy farms included in CADD agrees well with those permitted by CDFA.

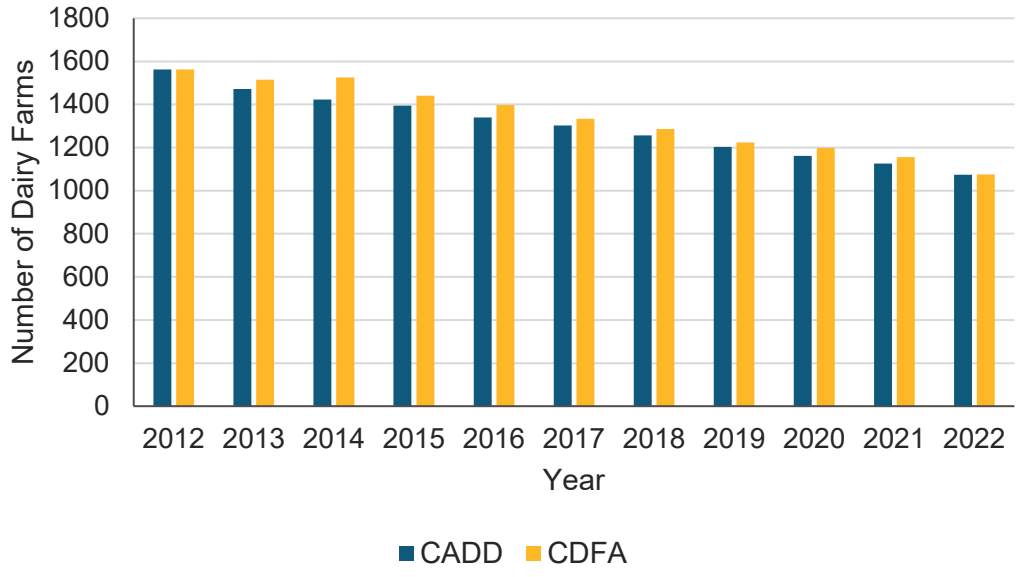

Figure 7. Number of cattle dairy farms included in CADD versus those permitted by CDFA.

A similar comparison was made between CADD and the Ag Census. The Ag Census is conducted every five years and relies on a voluntary survey of agricultural facilities that generate over $1,000 in agricultural product sales. The Ag Census provides statistics for numerous parameters, including cattle herd size, and these statistics are reported at the county, statewide, and national levels. As outlined in the Census methodology document,[18] adjustments are made to the data to address nonresponse, coverage, and misclassification errors. In the 2022 Ag Census, numerical adjustments were performed for 22.6% of the California dairy farms.[19] In addition, the Ag Census includes not only commercial

[17] *https://leginfo.legislature.ca.gov/faces/codes_displayexpandedbranch.xhtml?tocCode=FAC&division=15.&title=&part=&chapter=&article=&nodetreepath=18*

[18] *https://www.nass.usda.gov/Publications/AgCensus/2022/Full_Report/Volume_1,_Chapter_1_US/usappxa.pdf*

[19] Table A from *https://www.nass.usda.gov/Publications/AgCensus/2022/Full_Report/Volume_1,_Chapter_1_State_Level/California/cav1.pdf*

dairies but also farms with cows for home use. The contribution of these small-scale operations to the total population is negligible (For example, in 2022, total milk cows in farms with fewer than 10 mature cow herd size represented 0.03% of the total population), and CADD or other government agencies may not cover all of them. Staff analysis focused on dairy farms with a herd size of at least 10 mature cows to exclude farms that are most likely linked with cows for home use.

Figure 8 indicates that for all the years that both CADD and the Ag Census are available, CADD encompasses a larger number of dairy farms, demonstrating a higher level of coverage compared to the Ag Census. In 2022, the number of dairies (with a herd size of at least 10 mature cows) included in CADD was 25% higher than those captured in the Ag Census. These comparisons confirm that CADD's coverage of California dairies is more comprehensive than the Ag Census.

It should also be noted that while comparing animal populations in CADD with the Ag Census may appear useful, the coarse spatial resolution of the Ag Census limits the ability to identify the cause of any discrepancies.

As mentioned earlier, since some non-dairy cattle facilities like grazing operations are exempt from reporting under the Confined Bovine Feeding Operations General Order, CADD does not capture all non-dairy cattle facilities in the State.

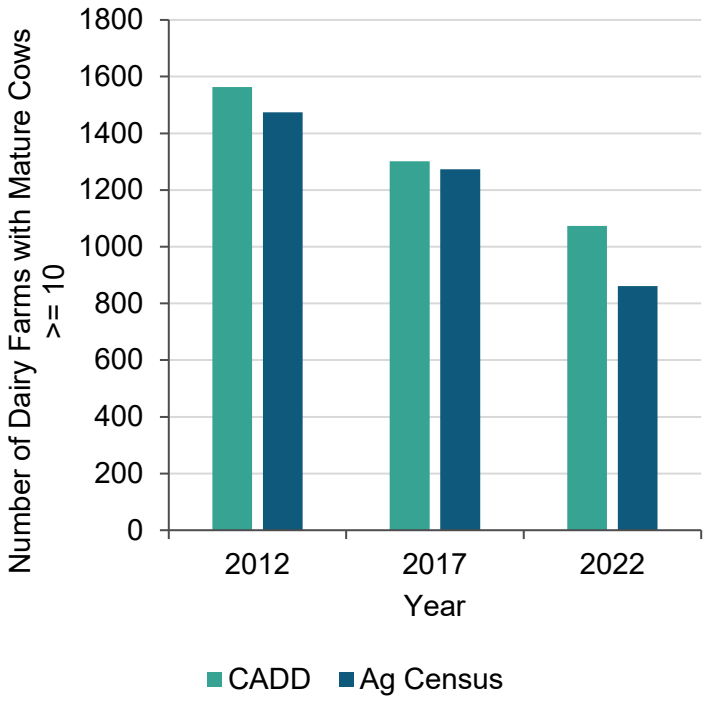

Figure 8. Number of dairy farms with at least 10 mature cows included in CADD and the Ag Census.

7 Summary

To improve the understanding of emissions trends in the dairy and livestock sector, CARB developed CADD using data from Federal, State, and Local agencies. The database provides facility-level information on cattle herd sizes and anaerobic digesters at California dairies from 2012 to 2022. Built from publicly accessible documents, CADD enhances CARB’s ability to track trends in the state's dairy

and livestock sector. As part of this effort, CARB staff digitized over 12,000 documents, many of which were paper copies. While Python scripts were used to improve efficiency, a substantial portion of records required manual digitization. All data in CADD underwent a rigorous QA/QC process to ensure accuracy.

An analysis of CADD's coverage revealed a close agreement with the number of dairies permitted by CDFA. Compared to the Ag Census, CADD offers improved frequency, spatial resolution, reliability, and comprehensiveness for dairies in California. With its high level of detail, CADD serves as a valuable resource for answering questions that were previously difficult to address.

8 Appendix A – Reference Code for Herd Size Data

Below is the description of the **MilkCowsHerdSizeRefCode** in CADD:

1: Milk cow herd size is based on an annual report submitted to the Regional Water Quality Control Boards.

2: No annual report is found. The facility is deemed not to be an active milking operation since

- **2a:** The Water Board records, including inspection reports and notes in CIWQS regulation information, indicated that the dairy was not an active milking operation during this year.
- **2b:** Notes in annual reports for other years or the notice of intent indicated that the dairy was not an active milking operation during this year.
- **2c:** Annual report(s) for other years with the same owner/operator reported zero milk cows.
- **2d:** Google Earth images indicated no animal or no milking parlor during this year or later years, while there was no annual report or inspection report with nonzero milk cows prior to this year for this facility with the same owner/operator.
- **2e:** Google Earth images indicated no animal or no milking parlor during the year or prior years while there was no annual report or inspection report with nonzero milk cows after this year for this facility with the same owner/operator.
- **2f:** Air permit for this facility with the same owner/operator indicated zero maximum milk cows.

3: No annual report is found but since none of the criteria under item 2 held, the facility is deemed to be an active milking operation and

- **3a:** The milk cow herd size was determined using herd sizes in annual reports for other years with the same owner/operator.
- **3b:** The milk cow herd size was determined using herd sizes in inspection reports for the same year or other years with the same owner/operator.
- **3c:** The milk cow herd size was determined using herd sizes in air permits for the same year or other years with the same owner/operator.
- **3d:** The milk cow herd size was determined using herd sizes in waste management plans for the same year or other years with the same owner/operator.
- **3e:** The milk cow herd size was determined using herd sizes in nutrient management plans for the same year or other years with the same owner/operator.
- **3f:** The milk cow herd size was determined using herd sizes in notice of intent for the same year or other years with the same owner/operator.
- **3g:** The milk cow herd size was set to zero due to the lack of herd size data in all the data sources.

Below is the description of the **NonMilkingCattleHerdSizeRefCode** in CADD:

1: Non-milking cattle herd size is based on an annual report submitted to the Regional Water Quality Control Boards.

2: No annual report is found. The facility is deemed to not house non-milking cattle since

2a: The Water Board records, including inspection reports and notes in CIWQS regulation information, indicated that the facility had no cattle during this year.

2b: Notes in annual reports for other years or the notice of intent indicated that the facility had no cattle during this year.

2c: Annual report(s) for other years with the same owner/operator reported zero animals.

2d: Google Earth images indicated no animal during this year or later years, while there was no annual report or inspection report with nonzero cattle herd size prior to this year for this facility with the same owner/operator.

2e: Google Earth images indicated no animal for this facility during this year or prior years, while there was no annual report or inspection report with nonzero cattle herd size after this year for this facility with the same owner/operator.

3: No annual report is found, but since none of the criteria under item 2 held, the facility is deemed to house non-milking cattle and

3a: The non-milking cattle herd size was determined using herd sizes in annual reports for other years with the same owner/operator.

3b: The non-milking cattle herd size was determined using herd sizes in inspection reports for the same year or other years with the same owner/operator.

3c: The non-milking cattle herd size was determined using herd sizes in air permits for the same year or other years with the same owner/operator.

3d: The non-milking cattle herd size was determined using herd sizes in waste management plans for the same year or other years with the same owner/operator.

3e: The non-milking cattle herd size was determined using herd sizes in nutrient management plans for the same year or other years with the same owner/operator.

3f: The non-milking cattle herd size was determined using herd sizes in notice of intent for the same year or other years with the same owner/operator.

3g: The non-milking cattle herd size was set to zero due to the lack of herd size data in all the data sources.

9 Appendix B – Regional Water Quality Control Boards' Orders Impacting Cattle Facilities

Table S1. Regional Water Quality Control Boards' Orders governing dairy and bovine feeding operations, along with their exemptions (as of March 2023).

Regional Water Quality Control Board Office	Dairy (D) or Bovine (B)	Order #	Order Title	Exemptions
1	D	R1-2019-0001	General Waste Discharge Requirements for Dairies Within the North Coast Region	<25 dairy cows
2	D & B	R2-2016-0031	General Waste Discharge Requirements for Confined Animal Facilities Within the San Francisco Bay Region	Notice of Non-Applicability: 1. The number of animals within confined areas is small in relation to the size of the facility and poses no potential for adverse water quality impact 2. grazing on pasture lands 3. Animals rarely confined, or 4. Stormwater does not contact manure, and all waste is disposed offsite.
3	D		Individual waste discharge requirements	NA
5	D	R5-2013-0122	Reissued Waste Discharge Requirements General Order for Existing Milk Cow Dairies	Coverage: "existing milk cow dairies": dairies that were operating as of 17 October 2005, filed a complete Report of Waste Discharge

				(ROWD) in response to the 2005 ROWD Request Letter and have not expanded since 17 October 2005.
5	D	R5-2010-130	Waste Discharge Requirements General Order for Dairies with Manure Anaerobic Digester or Co-Digester Facilities	None
5	B	R5-2017-0058	Waste Discharge Requirements General Order for Confined Bovine Feeding Operations	Operations where cattle representing fewer than 6 Animal Units (AU) are confined.
5	D		Individual waste discharge requirements	NA
6	D & B		Individual waste discharge requirements	NA
7	D & B	R7-2021-0029	General Waste Discharge Requirements for Concentrated Animal Feeding Operations Within the Colorado River Basin Region	None
8	D & B	R8-2018-0001	General Waste Discharge Requirements for Concentrated Animal Feeding Operations (Dairies and Related Facilities) Within the Santa Ana Region	Dairies with animal population < 20 (dry and milking) or heifer, calf, or cattle ranches with herd size < 50
9	D & B	R9-2008-0130	Waste Discharge Requirements for Existing Dairy Animal Feeding Operations in the San Diego Region	None

Table S2. Range of reported years for 2012–2022 for each Regional Water Quality Control Board.

Region	Reported Years										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	X	X	X	X	X	X	X	X	X	X	X
2					X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X	X	X
5 (Dairy GO)	X	X	X	X	X	X	X	X	X	X	X
5 (Bovine Feeding Operation GO)								X	X	X	X
6		X	X	X	X	X	X	X	X	X	X
7			X	X	X		X	X	X	X	X
8	X	X	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	X	X	X

GO: General Order

Table S3. Reporting format, type of documents, and reported animal types provided by each Regional Water Quality Control Board.

Region	Reporting Format	PDF/SS	Reported Animal Types												
			MC	DC	OH	YH	OC	YC	BC	MatC	H	C	NMatC	NMC	TC
1	R1	SS								X			X		
2	R2	PDF & SS								X			X		
3	R3	PDF	X											X	
5	Merced Tool	PDF	X	X	X	X	X	X							
5	Innovative Ag – Dairy GO	PDF	X	X	X	X	X	X							
5	Livingston – Dairy GO	PDF	X	X	X	X						X			
5	Provost & Pritchard	PDF	X	X	X	X	X	X							
5	Bovine Format - Confined Bovine Feeding Operations Only	PDF		X	X	X	X	X	X						
5	Innovative Ag – Confined Bovine Feeding Operations GO	PDF		X	X	X	X	X	X						
5	Livingston – Confined Bovine Feeding Operations GO	PDF		X	X	X			X			X			
6	R6	PDF								X	X	X			
7	R7	SS													X
8	R8	PDF & SS	X	X					X		X	X			
9	R9	PDF	X	X							X	X			

SS: Spreadsheet, MC: Milk Cows, DC: Dry Cows, OH: Old Heifers, YH: Young Heifers, OC: Old Calves, YC: Young Calves, BC: Beef Cows, MatC: Mature Cows, H: Heifers, C: Calves, NMatC: Non-Mature Cows, NMC: Non-Milk Cows, TC: Total Cows, GO: General Order.

10 Appendix C – Identifying Potentially Operational Facilities

CIWQS database with the following filters was used to generate a list of facilities that were suspected to be operational at any time since 2012:

1. CAFO Type: “Dairies”, “Feedlots”, “Blank”, “NA”, or “Other”,
2. CAFO SUBTYPE: “Calf feedlots”, “Cattle or cow/calf pairs”, “Finishing Yards/Auction Yards”, “Heifers (non dairy affiliated)”, “Mature dairy cattle”, “NA”, or “Blank”,
3. Place Type: “Animal Feeding”, or “Facility”,
4. SIC Desc 1: “Agricultural production- livestock”, “Beef Cattle Feedlots”, “Beef Cattle, Except Feedlots”, “Dairy Farms”, “NA”, or “Blank”, and
5. Termination Date: “Blank” or “≥ 2012”.

The filtered facilities were investigated using Google Earth images, and if animals existed in a facility at any time between 2012 and 2022, it was added to CADD.

11 Appendix D – Decision Tree to Address Missing Herd Size Data

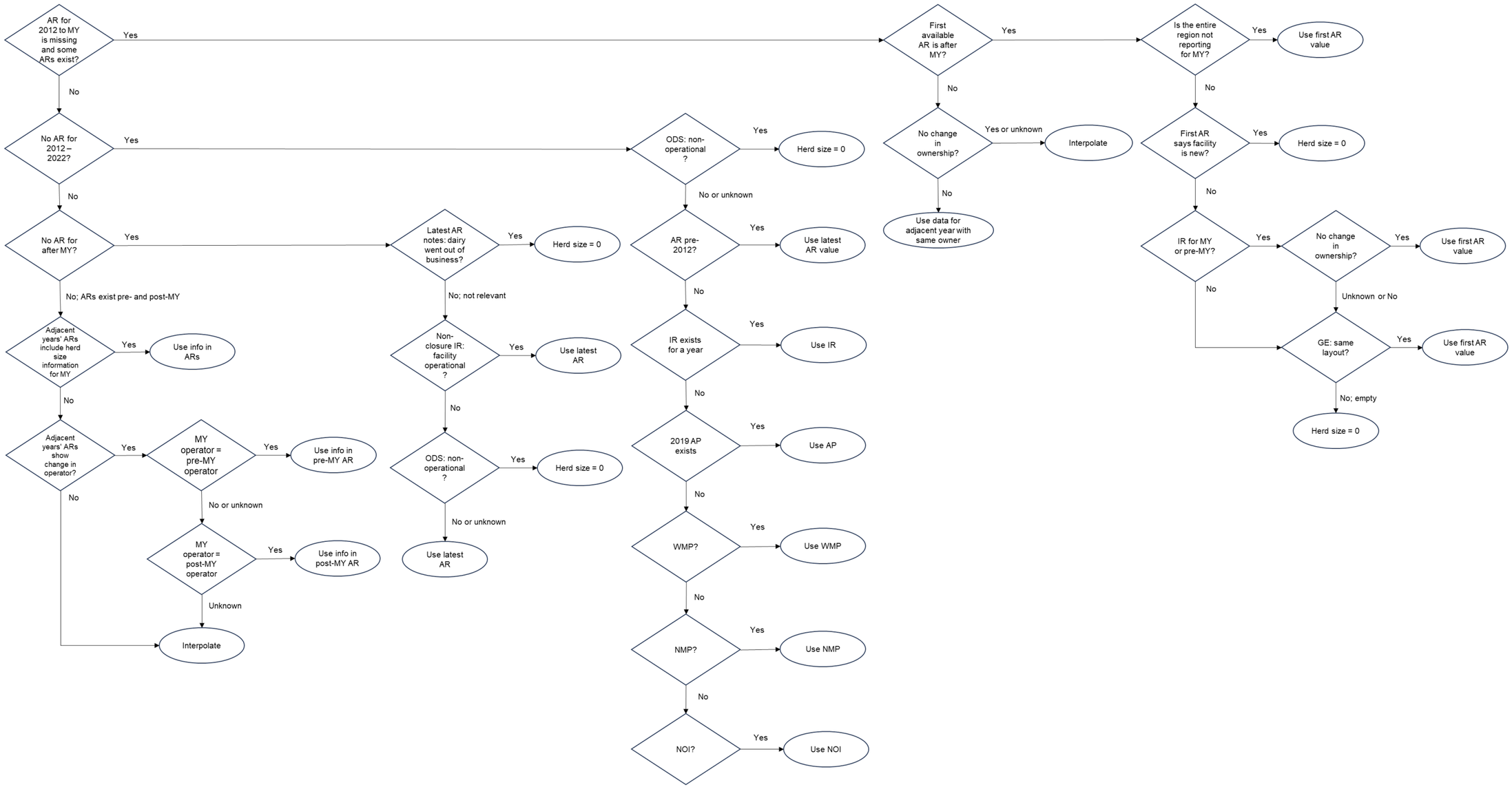

Figure S1. Decision tree to fill missing herd size data. For facilities with 2019 – 2022 data coverage in CADD, replace 2012 with 2019. In the context of addressing milk cow herd size data, an operational facility refers to a dairy with active milking operation while in the context of addressing non-milking cattle herd size data, an operational facility refers to a facility that houses cattle. AR: Annual Report, MY: Missing Year, IR: Inspection Report, GE: Google Earth, AP: Air Permit, WMP: Waste Management Plan, NMP: Nutrient Management Plan, NOI: Notice of Intent, ODS: Other Data Sources that are used to determine if a facility was inactive (refer to Appendix A for more information).