

# 2023 Zeroing Methane Emissions Datathon Quantitative Guide

Date: February 15, 2023

Revision 1 Updates: Any new updates will be indicated in red.

## 1. Official Kaggle Challenge

Due to the change in dataset, a new Kaggle competition had to be created. The official Kaggle site is located in the following URL: <http://www.kaggle.com/competitions/2023-ch4-emissions-datathon>

However, in order to join the challenge, participants must use the following URL:

<https://www.kaggle.com/t/e4e92e11b90e433a82ce34caaeb80190>

## 2. Quantitative Criteria: Feature Engineering

In this section of the competition you are tasked with building an image regression machine learning (or other) model that will be able to predict the methane emission rate (kg/hr) for twenty (20) given satellite images. These will be based on the actual methane emission rate associated with the satellite image, not the persistence adjusted rates. **Geotiffs will not be provided, only the test png files will be provided.** A few rules:

1. Teams are only limited to the data referenced by the organizers of the competition. This means that any dataset found in the DropBox (<http://tiny.cc/zerodata>) and any of the original rasters associated with the candidate\_id that are mentioned in following Excel files can be used:
  - a. permian\_plume\_list\_2019 Jeremy Zhao.csv
  - b. plume\_attribution\_2019-2022 Jeremy Zhao.csv

As an example, if “ang20190922t192642-2” is mentioned in the list, then the original associated rasters for that associated candidate\_id may be used as part of your prediction. These original rasters can be downloaded from the following links:

<https://zenodo.org/record/5610307>

<https://zenodo.org/record/5606120>

<https://zenodo.org/record/7072824>

2. Top 3 performing teams (based on private leaderboard) will be asked to open-source their code and model to ensure that other participants and broader Calgary Data Science Community can learn from their solutions.
3. No AutoML services.
4. The scoring will be root mean squared error.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

5. Teams are advised to generate their own png files from any allowed rasters from the associated files as mentioned above.

### 3. Quantitative Team Submission

The Quantitative team submission is done through Kaggle:

<http://www.kaggle.com/competitions/2023-ch4-emissions-datathon> following these rules:

- Please register one account per team for submissions.
- There are 5 submissions per day allowed on the public scoring section.
- Each team will also be allowed 2 eligible submissions for the final private leaderboard. Users can hand-select the eligible submissions, or will otherwise default to the best public scoring submissions.

Submissions are evaluated in real time and will be received until **Saturday, March 4, 2023 at 11:59pm** driving leaderboard ranking. The winning team for the Quantitative part of the competition will be announced at the Closing Ceremony on Tuesday, March 7, 2023. Please read all the rules found on the Kaggle page.

### 4. Data

The template on how to submit, and the test files for your machine learning model can be found under the data section of the Kaggle competition site.

### 5. Team Formation Note

Teams are not free to form and re-form throughout the competition. The team used for the quantitative portion of the challenge should be the same team recognized at the time of the final video submission.

### 6. Rule Changes

The organizers will continue to exercise the ability to change the rules if deemed necessary to ensure smooth operation of the Kaggle competition. The organizers also reserve the right to request the a reproducible code upon request to ensure the integrity of the competition, but this will only be exercised in under extreme circumstances.