mad

methane anomaly detector

“We’re MAD about methane”

The Team

Alyssa Augsburger
Senior Member of Technical Staff

Jaclyn Andrews
Director

John Lee
Mechanical Engineer

Karthik Rameshbabu
Senior Software Engineer

Sanjay Saravanan
Software Development Engineer

“

The Environmental Protection Agency estimates that

methane causes about

25x

more damage than carbon dioxide

to the atmosphere over a 100 year period.

PHYS ORG

NOVEMBER 10, 2021

Methane's short lifespan presents golden opportunity to quickly address climate change

by Christina Procopiou, Lawrence Berkeley National Laboratory

FAST COMPANY

11-29-21 | 1:30 PM | WORLD CHANGING IDEAS

Stemming methane leaks could be key to slowing global warming quickly

npr

Biden joins global push to cut climate-warming methane emissions

Updated November 2, 2021 · 2:36 PM ET

The Problem

- Detecting methane leaks is not easy and not cheap
- No simple available resource for monitoring methane
- Methane-emitting facility owners need our help to detect methane leak warning signs

The Solution

- Machine learning anomaly detection
  - Growing data set
  - Capacity for other signals

- Data as a Product
  - Clean data access

Data

Data Pipeline

Methane (Sentinel 5P)

- Accessibility
- Credibility
- Frequency

Weather (ERA5)

- Data Quality
- Relevance

Data Grouping (Climate Zone)

- Address Lack of Data
- Data Grouping

Process and Merge

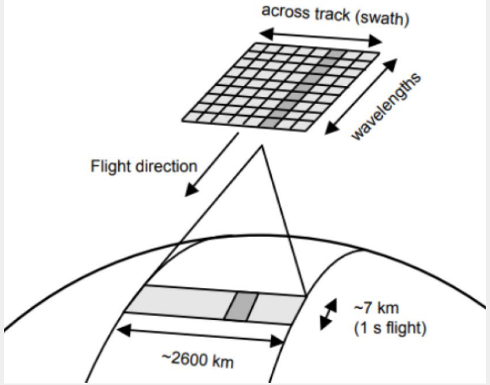
- Model Ready
- Updated Real-Time

Methane

- **Sentinel 5P:**
  - European Space Agency satellite that monitors atmospheric conditions
  - **Credible**
  - Publicly **Accessible**
  - Daily **Frequency**
- **Data:**
  - Methane concentrations (ppb)
  - 7km x 5.5km resolution
  - 90GB raw -> 34mb extracted

TROPOMI

sentinel-5p



Methane Data

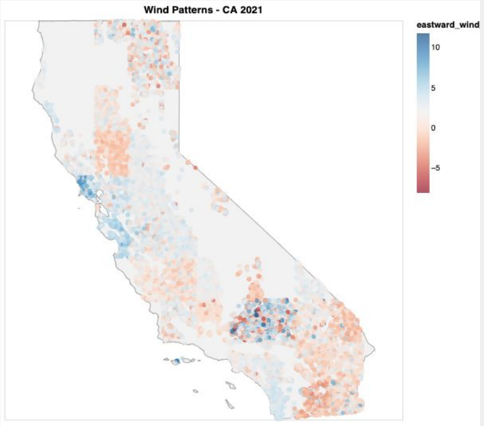
Weather Data

Data Grouping

Process and Merge

Weather

- **Why ERA5:**
  - Methane **data quality** influenced by weather
  - Methane concentrations can be **correlated** to weather
- **ERA5 Data:**
  - European Centre for Medium-Range Weather Forecasts
  - Hourly Frequency
  - 250GB raw -> 1.3GB extracted



Methane Data

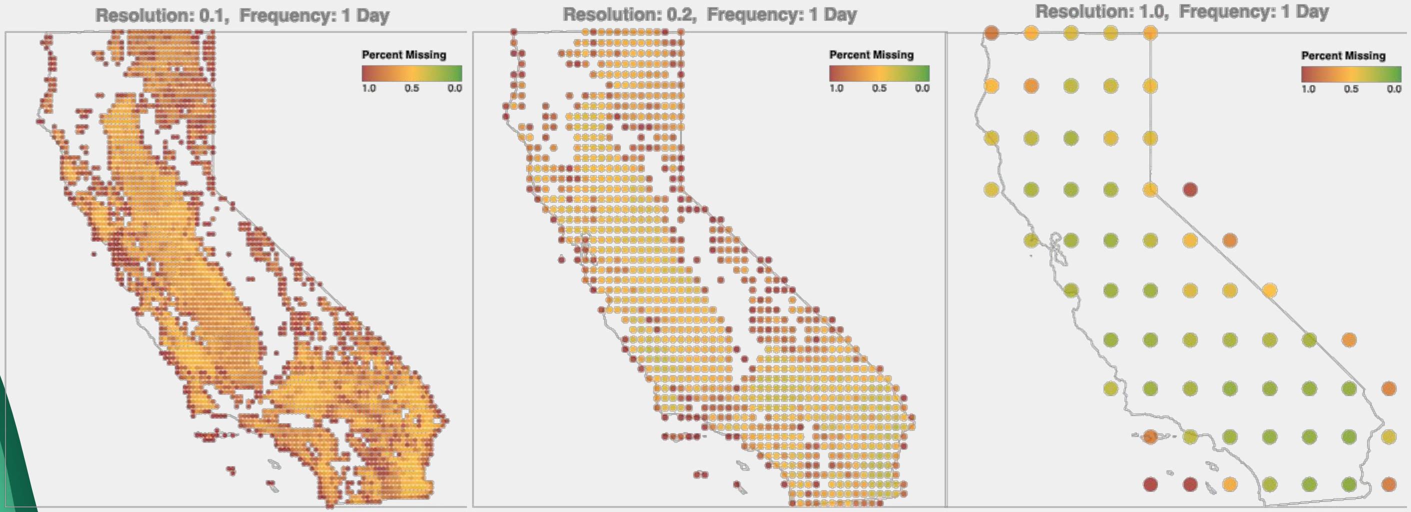
Weather Data

Data Grouping

Process and Merge

Data Grouping

Higher Resolution ← → Less Missing Data

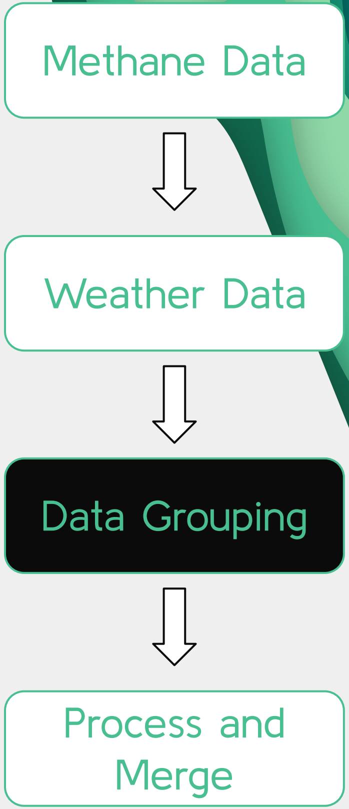


Variations: Rounded lat/lon of 0.1, 0.2, 1.0

Methane Data
↓
Weather Data
↓
Data Grouping
↓
Process and Merge

Data Grouping

- **Why Zones:**
    - Aggregated data **reduces data sparsity**
    - A method of **grouping**
- **California Climate Zones:**
    - 16 zones adopted by California
    - Varies by climate conditions (e.g. temperature, wind speed, solar radiation)



Process and Merge

Currently: 1,251,651 rows & 31 columns

time	methane	lat	lon	rn_lat_5	rn_lon_5	...	zone
2018-11-28	1879.700317	32.822247	-114.522209	33	-114.5	...	15
2018-11-28	1883.997314	32.726692	-114.938751	32.5	-115.0	...	15
...	...	...	...	...	...	...	...
2021-10-21	1896.328003	34.465900	-116.691612	34.5	-117.0	...	14
2021-10-21	1892.318359	34.465900	-116.691612	34.5	-116.5	...	14

Methane Data

Weather Data

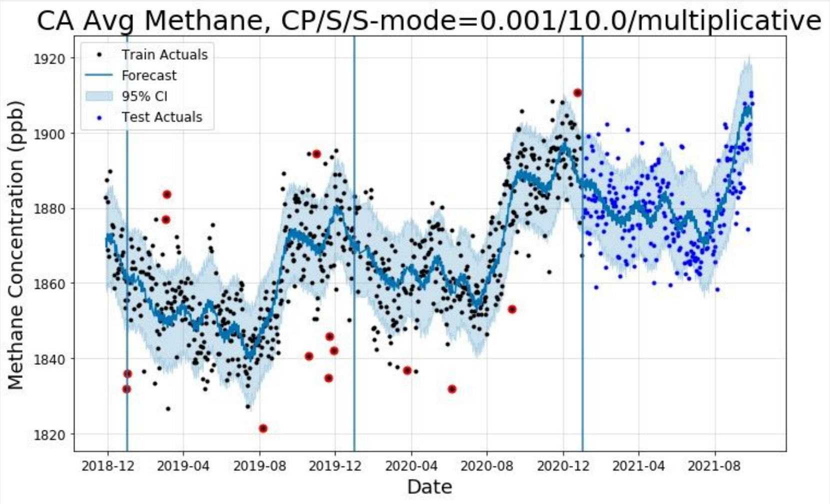
Data Grouping

Process and Merge

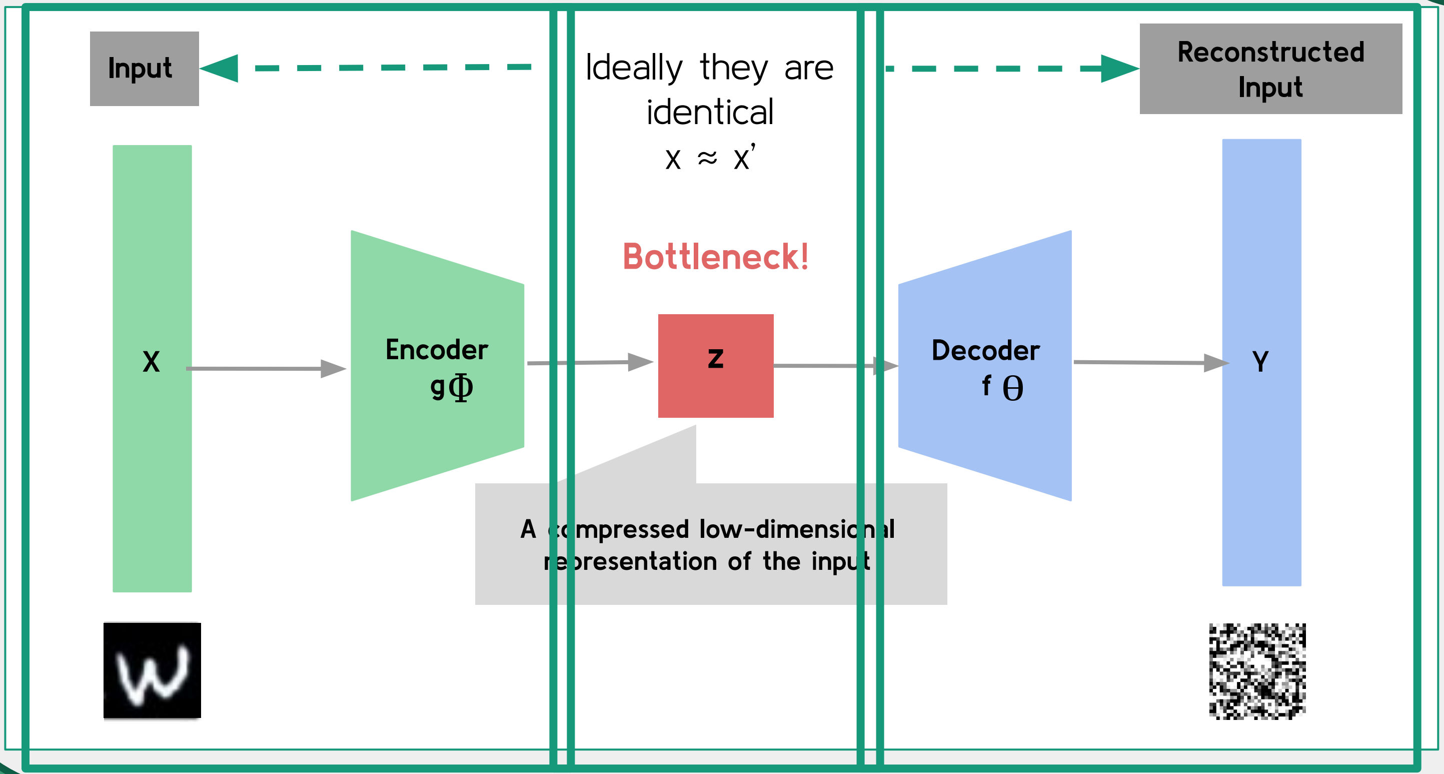
Models

Statistical Modeling

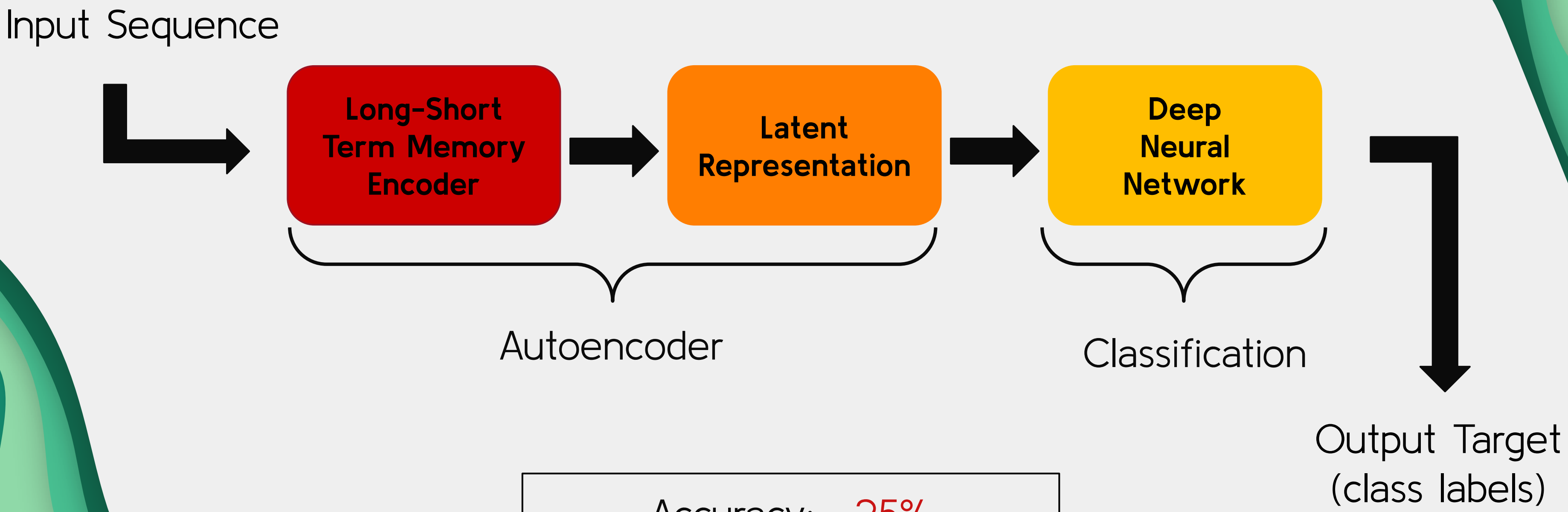
- Facebook Prophet
- Created baseline time series model using methane readings averaged across CA
- Only captures temporal aspect of the data
- **Limitation**
    - Could not incorporate additional regressors more effectively



Autoencoder Deep Learning

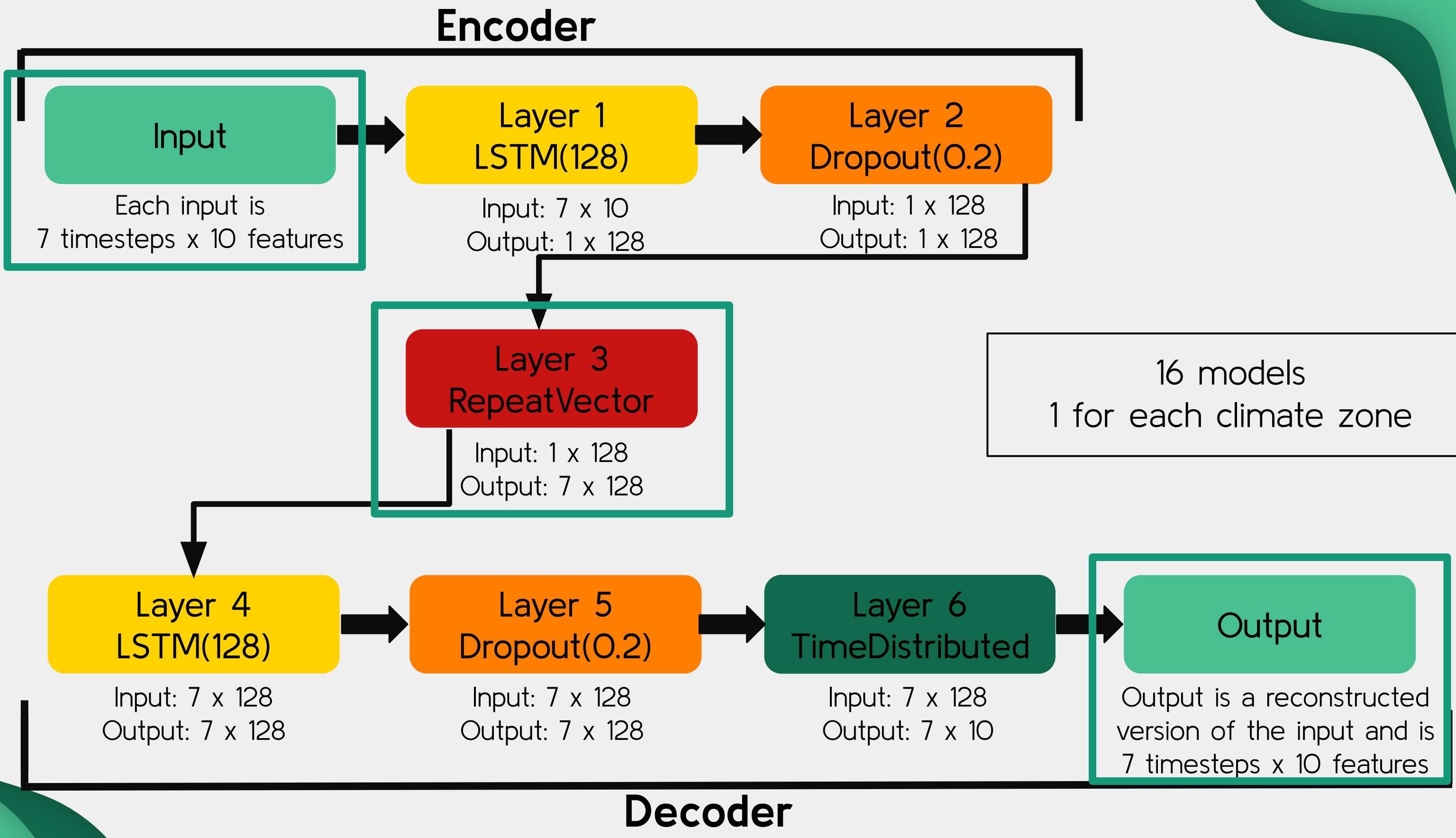


Autoencoder + Deep Neural Network
(Approach #1)



Accuracy: ~25%
Predicted: ~75% anomalies

Long Short Term Memory (LSTM)
(Final Autoencoder Framework)



16 models
1 for each climate zone

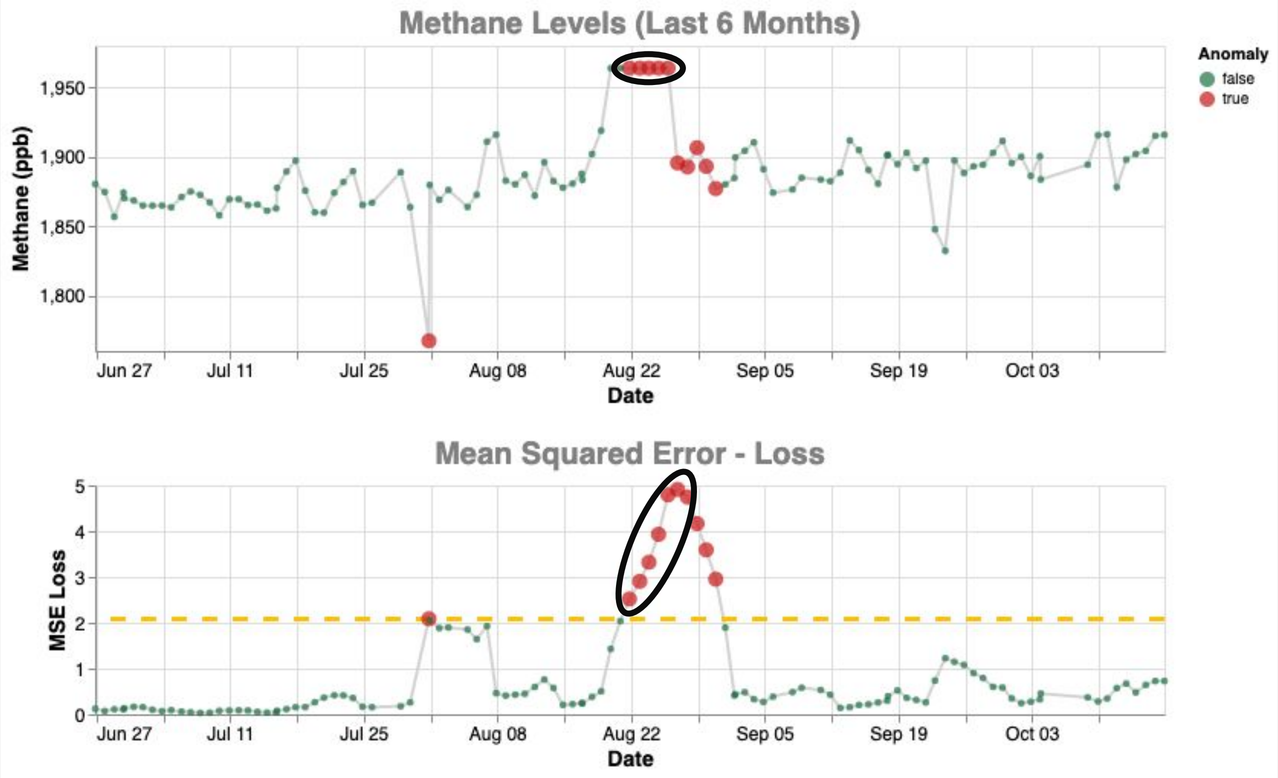
Model Tuning

- Hyperparameter tuning
  - Number of nodes
  - Dropout rates
  - Optimizers
  - Time window size
  - Epochs
- Data tuning
  - Resolution
  - Frequency

Climate Zone 1 Results		
Data	**Nodes**	**Root Mean Square Error (RMSE)**
Train	64	0.7119
Train	128	0.6862
Validation	64	0.8206
Validation	128	0.7833

4.5%

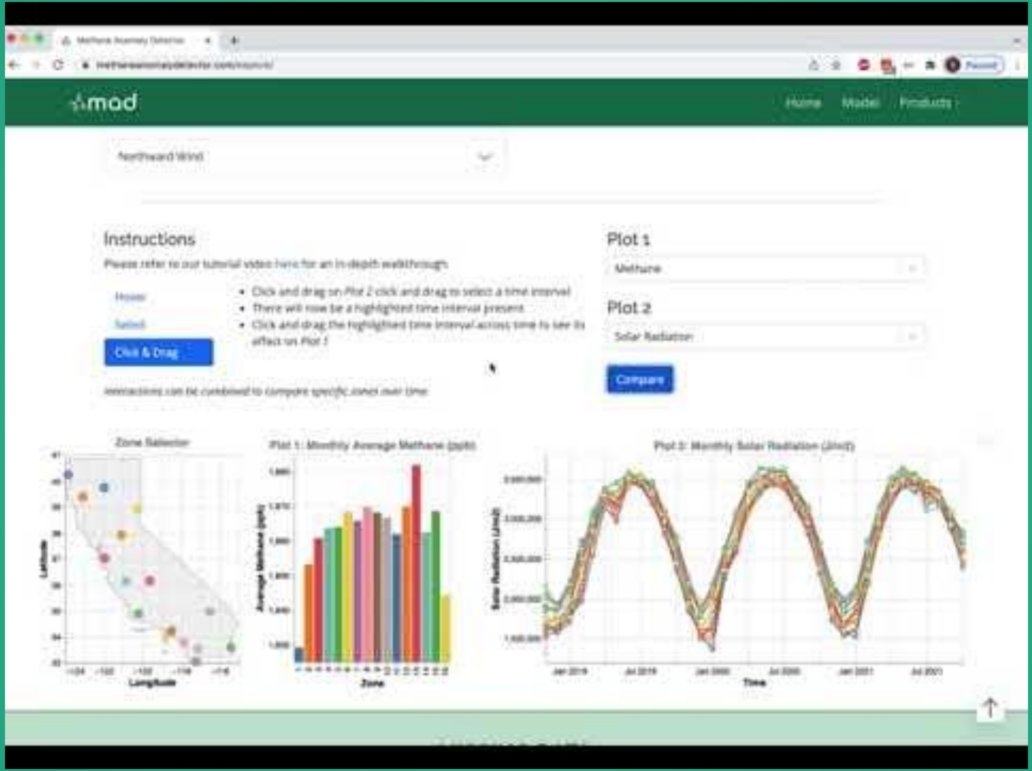
Synthetic Data Analysis



Single Actual Data Point

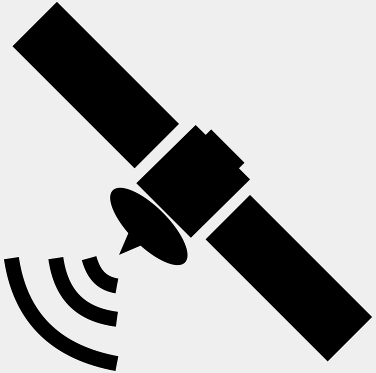
Series Synthetic Data

Product Demo

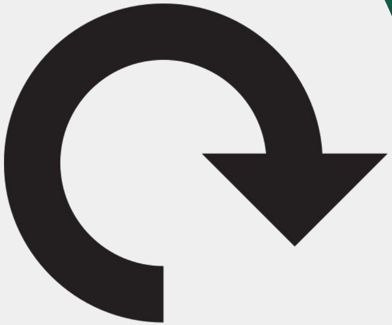

mad
Home
Model
Products
Northward Wind
Instructions
Please refer to our tutorial video here for an in-depth walkthrough
Hover
Select
Click & Drag
Click and drag on Plot 2 click and drag to select a time interval
There will now be a highlighted time interval present
Click and drag the highlighted time interval across time to see its effect on Plot 1
Interactions can be combined to compare specific zones over time
Plot 1
Methane
Plot 2
Solar Radiation
Compare
Zone Selector
Plot 1: Monthly Average Methane (ppb)
Plot 2: Monthly Solar Radiation
Longitude
Latitude
Zone
Time

Next Steps

Scale globally



Expand to different satellites



Retrain models with growing data

Thank You

- **Blair Jones,** Methane Enthusiast
- **Sam Schabacker,** Renewable Energy Manager, UC President's Office
- **Bill Collins,** Director of the Climate & Ecosystem Science Division at Lawrence Berkeley National Laboratory & Lead and Collaborating Author of the Fourth IPCC Assessment, for which the IPCC was awarded the *2007 Nobel Peace Prize*.
- **Iman Gohari,** Weather Forecasting Expert
- **Ed Dlugokencky,** NOAA Research Chemist
- **Jeff Zimmerman,** Environmental Engineering Manager, AECOM
- **Rick Malmstrom,** Executive Director-Sustainability Operations, Alexandria Real Estate
- **Cheryl Laskowski, Ryan Schauland, Jorn Herner,** California Air Resources Board (CARB)
- **Alberto Todeschini, Fred Nugen, Colorado Reed,** Faculty
- **Fellow classmates**

mad

Thanks!

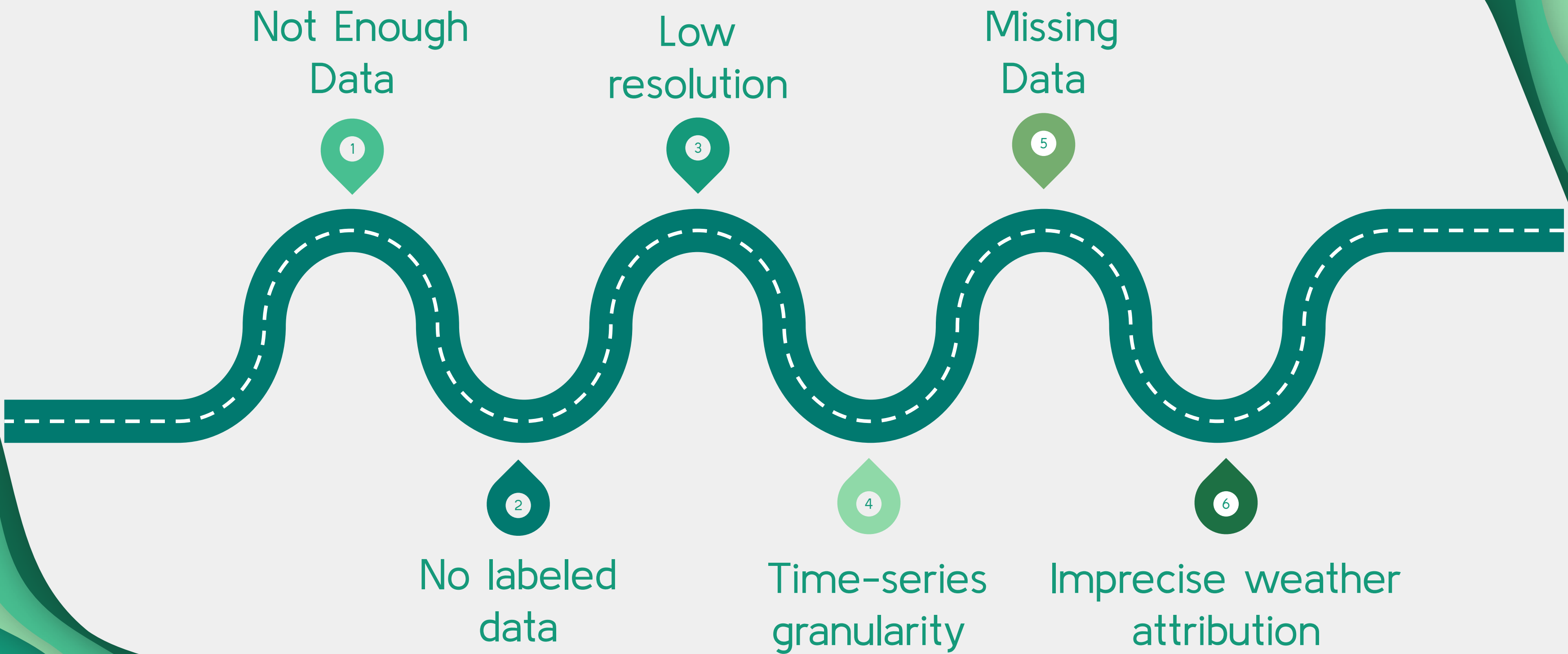
Any questions?

You can find us at:

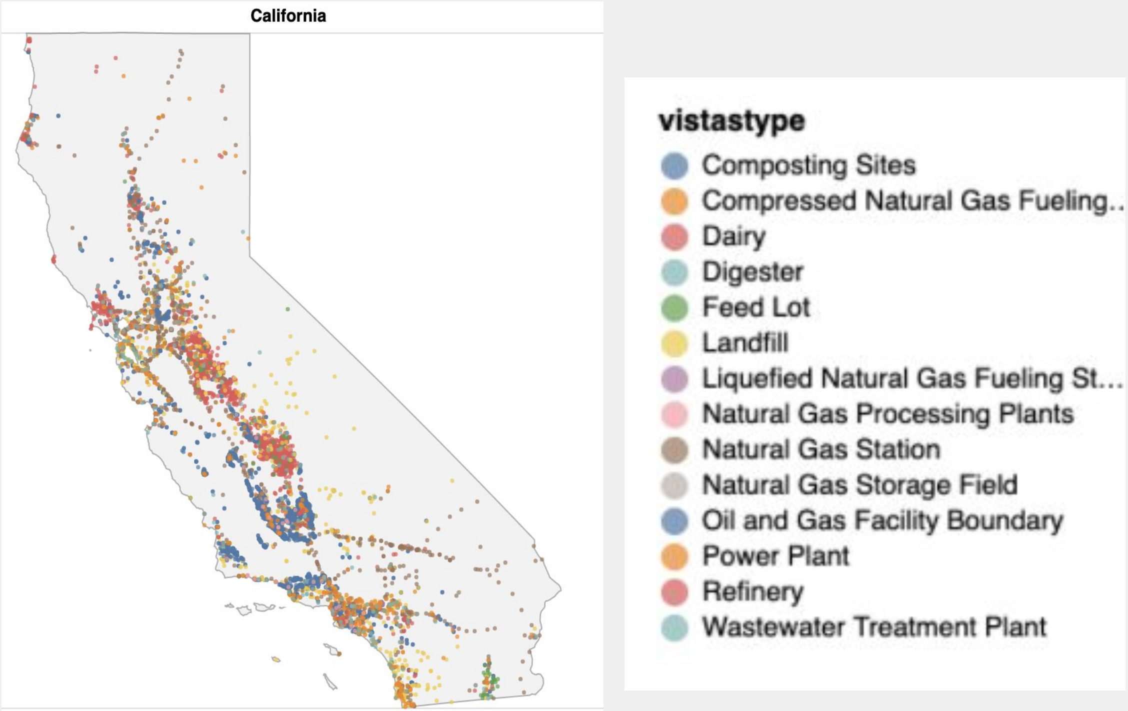
https://methaneanomalydetector.com/

Appendix

Appendix - Data Challenges



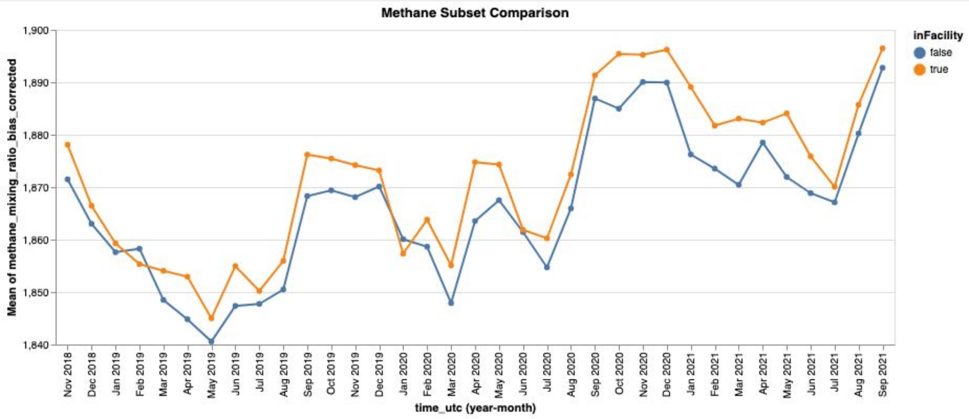
Appendix - Data - Vista-CA

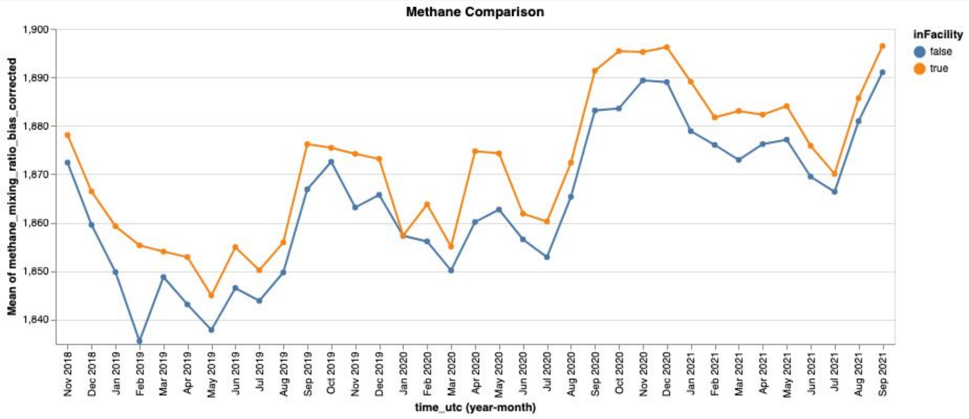


	area	vistastype
	mean	count
vistastype		
Composting Sites	0.001520	860
Compressed Natural Gas Fueling Station	0.000887	162
Dairy	0.000000	1715
Digester	0.020325	33
Feed Lot	0.000000	72
Landfill	1.374963	714
Liquefied Natural Gas Fueling Station	0.000467	46
Natural Gas Processing Plants	0.356001	26
Natural Gas Station	0.000000	1120
Natural Gas Storage Field	25.234850	12
Oil and Gas Facility Boundary	0.007556	3356
Oil and Gas Well	0.000000	225766
Power Plant	0.064109	433
Refinery	2.182393	26
Wastewater Treatment Plant	0.670160	149

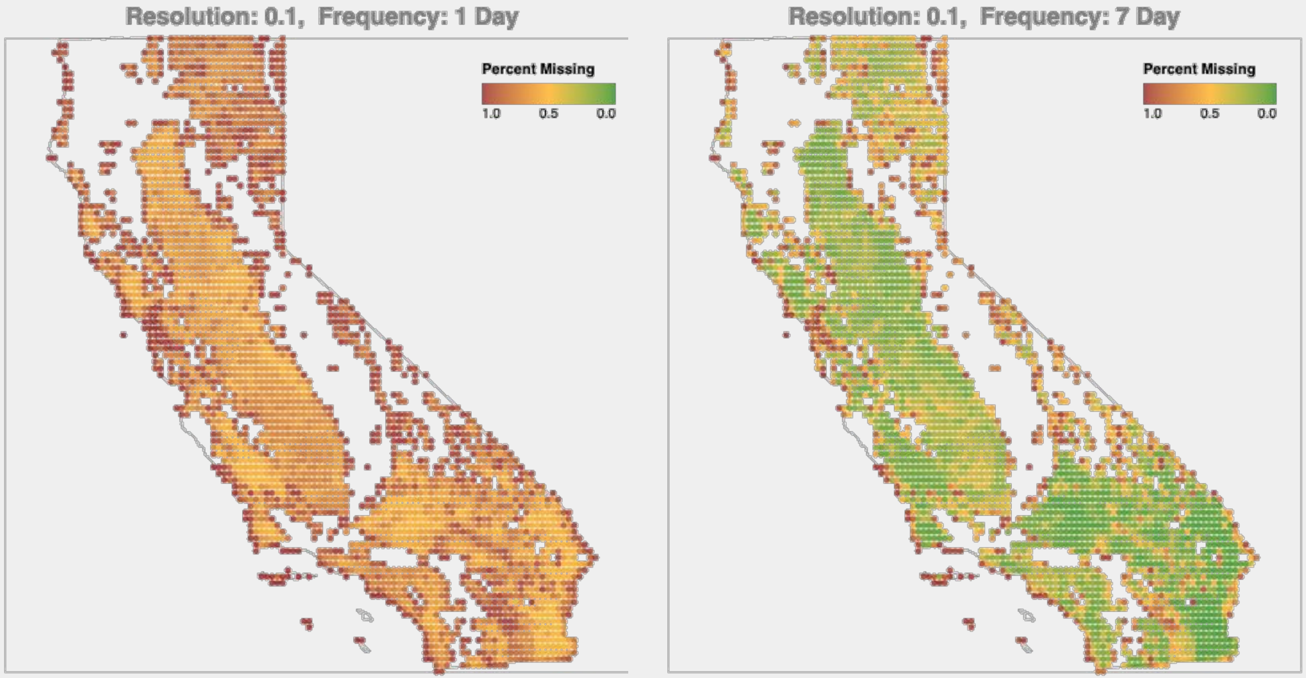
```
inFacility
False    1215953
True        3381
```

Appendix - Data - Vista-CA





Appendix - Data Grouping



Variation: Utilize Frequencies of 1D, 3D, 5D, 7D, 10D

Methane Data

⇩

Weather Data

⇩

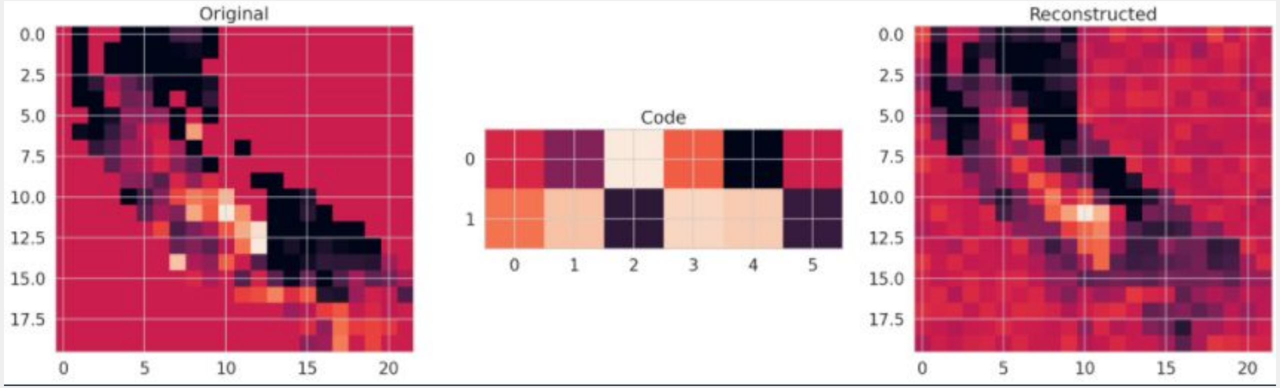
Data Grouping

⇩

Process and Merge

Appendix - Image Representation

- Represented methane readings as an image to capture spatial aspect. Each pixel represents methane reading at a location.
- Used a convolutional neural network (CNN) autoencoder to reconstruct images and measure loss
- Infeasible due to lack of data and lack of precision
- MSE = 0.1245



Appendix
Autoencoder Time Series Training

```
print(train[['mmrbc_i']].shape)
train[['mmrbc_i']].head(10)
```

(760, 1)

	mmrbc_i
time_utc	
2018-12-03	0.388968
2018-12-04	0.332599
2018-12-05	0.276231
2018-12-06	0.219866
2018-12-07	0.163498
2018-12-08	0.107129
2018-12-09	0.050760
2018-12-10	-0.005604
2018-12-11	-0.061973
2018-12-12	-0.118341

Sequence Windowing

```
trainX.shape:  (753, 7, 1) trainY.shape:  (753,)
testX.shape:  (266, 7, 1) testY.shape:  (266,)
```

Window #0

```
print("X")
print(trainX[0])
print()
print("Y")
print(trainY[0])
```

```
X
[[0.38896778]
 [0.3325992 ]
 [0.27623063]
 [0.2198662 ]
 [0.16349763]
 [0.10712906]
 [0.05076048]]

Y
-0.0056039393
```

Window #1

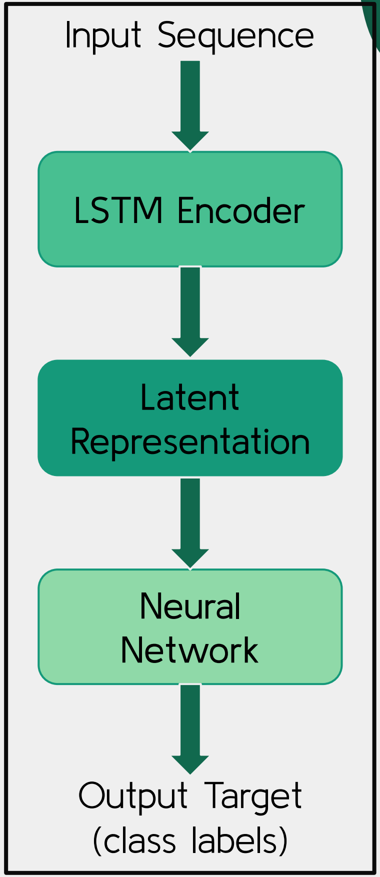
```
print("X")
print(trainX[1])
print()
print("Y")
print(trainY[1])
```

```
X
[[ 0.3325992 ]
 [ 0.27623063]
 [ 0.2198662 ]
 [ 0.16349763]
 [ 0.10712906]
 [ 0.05076048]
 [-0.00560394]]

Y
-0.061972514
```

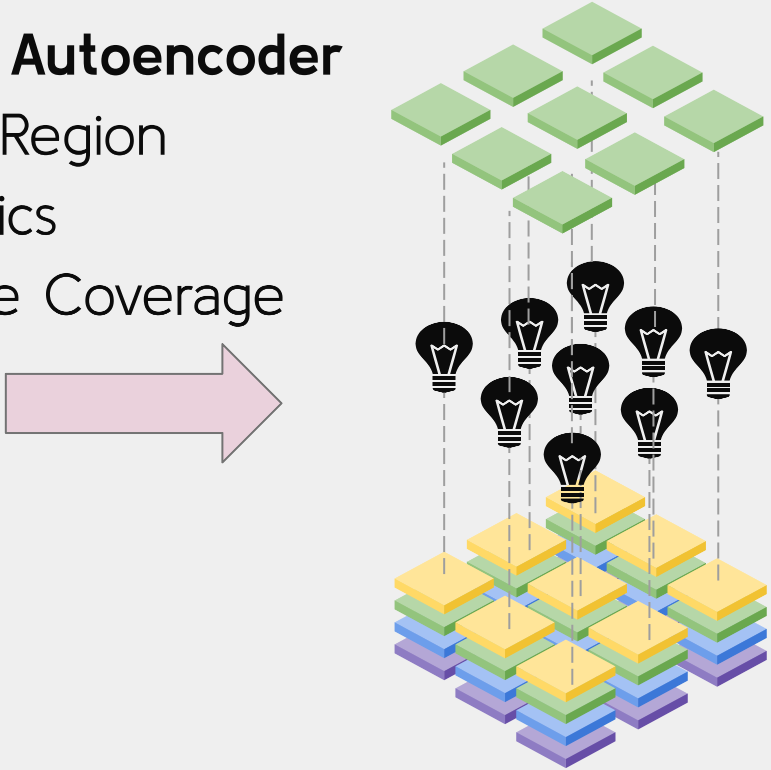
Appendix - LSTM-DNN

- Long Short Term Memory (LSTM) = capture temporal aspect
- Deep Neural Network (DNN) = capture spatial aspect
- Created a multiclass classification problem
- Achieved 25% accuracy
- Infeasible due to lack of data



Appendix - Final Approach

- **Multivariate LSTM Autoencoder**
- Model Trained per Region
- Multivariate Heuristics
- Time Series Feature Coverage
    - Methane
    - Weather
    - Q[illegible]
    - [illegible]
    - C[illegible] Count



Appendix - Feature Importance Example

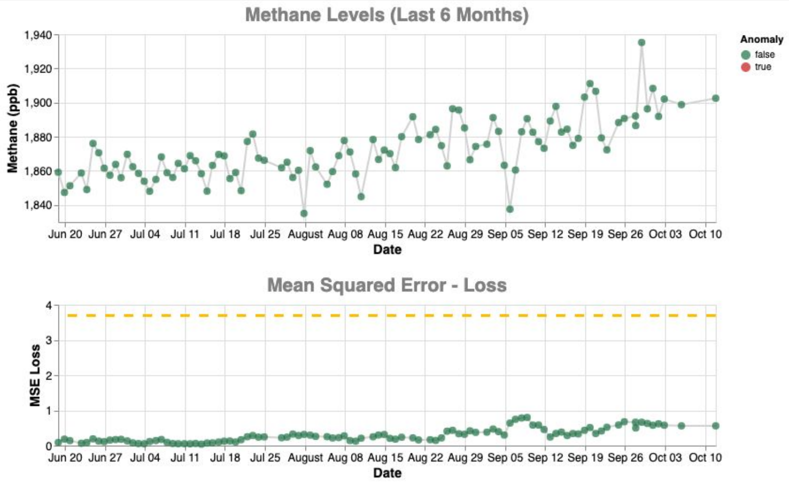
Model Name	All Features	Reading Count	QA Val Mean	QA Val Mode	Air Pressure at Mean Sea Level Mean	Air Temperature at 2 Meters Mean	Air Temperature at 2 Meters 1 Hour Maximum Mean	Air Temperature at 2 Meters 1 Hour Minimum Mean	Dew Point Termperature at 2 Meters Mean	Eastward Wind at 100 Meters Mean	Eastward Wind at 10 Meters Mean	Integral WRT Time of Surface Direct Downwelling Shortwave Flux in Air 1 Hour Accumulation Mean	LWE Thickness of Surface Snow Amount Mean	Northward Wind at 100 Meters Mean	Northward Wind at 10 Meters Mean	Precipitation Amount 1 Hour Accumulation Mean	Snow Density Mean	Surface Air Pressure Mean	Minimum RMSE
TRAIN RMSE																			
Zone 1	0.7061	0.6391	0.6987	0.7580	0.6698	0.6348	0.6219	0.6206	0.6781	0.7119	0.7210	0.5776	0.7529	0.6908	0.6908	0.7533	0.6554	0.7159	0.5776
VALIDATION RMSE																			
Zone 1	0.9118	0.5903	0.7219	0.7713	0.6474	0.5619	0.5589	0.5602	0.5683	0.6795	0.6721	0.5451	0.5334	0.6808	0.6933	1.0720	0.9360	0.6373	0.5334
METHANE RMSE ONLY WITH ONLY FEATURES ADDED PER COLUMN																			
Zone 1	0.7094	0.3207	0.3839	0.4851	0.4982	0.4952	0.4970	0.5027	0.5194	0.5194	0.5006	0.5452	0.5472	0.4184	0.4511	0.5326	0.5908	0.3717	0.3207

Appendix - RMSE Results

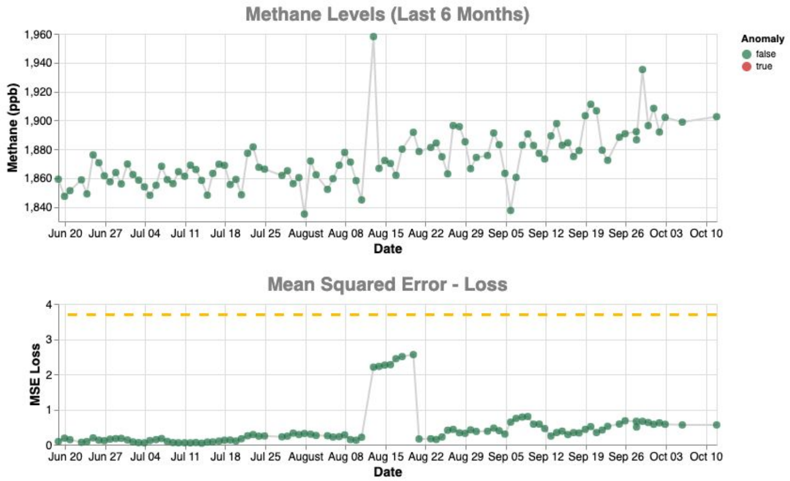
Zone	Train RMSE	Validation RMSE
1	0.7148	0.6035
2	0.6778	0.6195
3	0.6695	0.5930
4	0.6549	0.5816
5	0.6172	0.6612
6	0.6317	1.9263
7	0.6519	0.6765
8	0.6850	0.6332
9	0.6976	0.5981
10	0.6367	0.6937
11	0.6973	0.5853
12	0.6644	0.5887
13	0.7172	0.6370
14	0.6717	0.6095
15	0.6607	0.6729
16	0.7005	0.6226

Appendix - Santa Maria - Synthetic Test - Single Anomaly

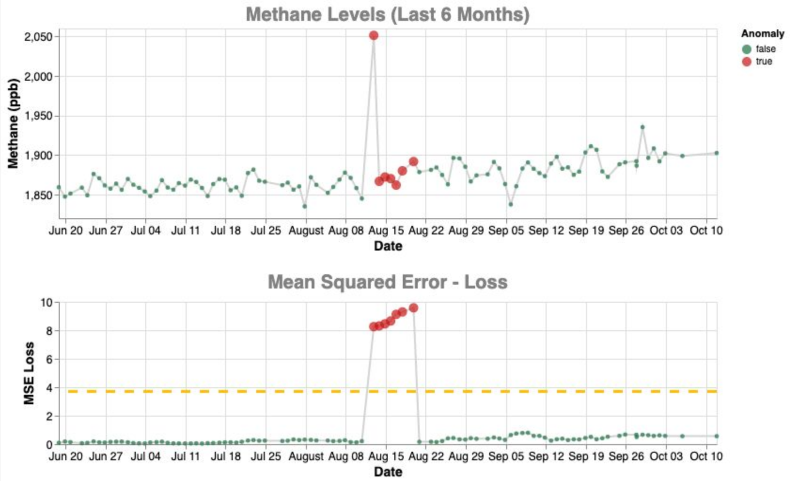
True



5% + mean

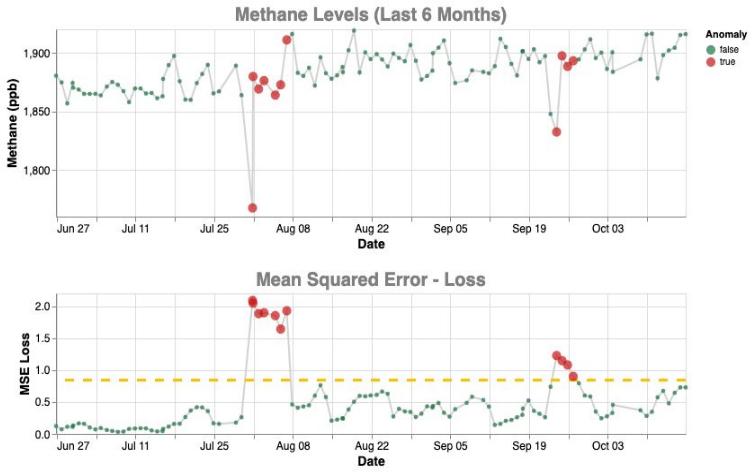


10% + mean

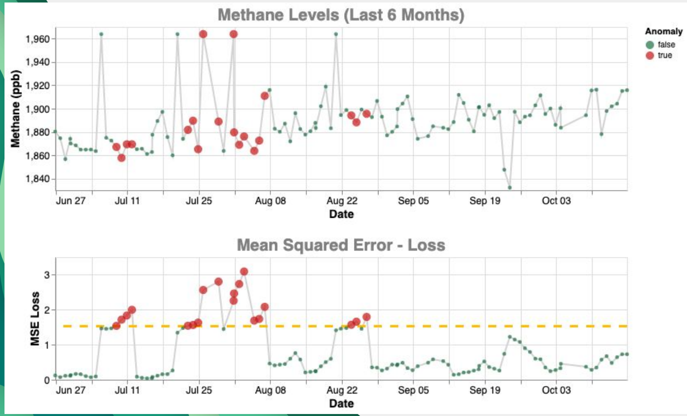


Appendix - Neal Road - Synthetic Test - Random Sequence Anomalies

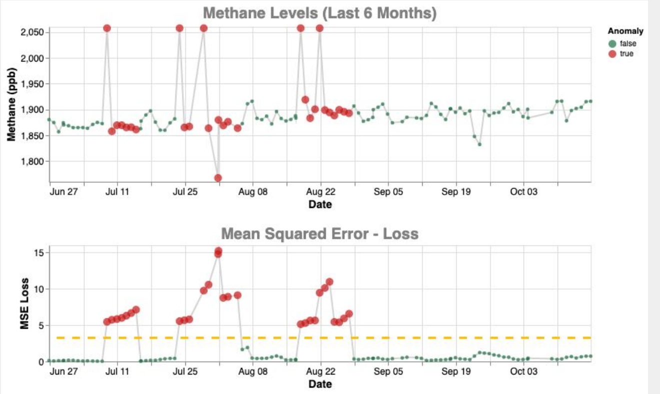
True



5% + mean

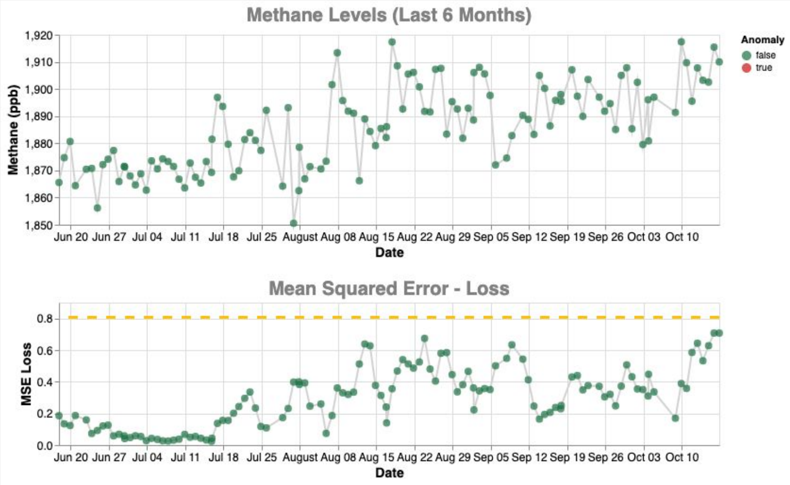


10% + mean

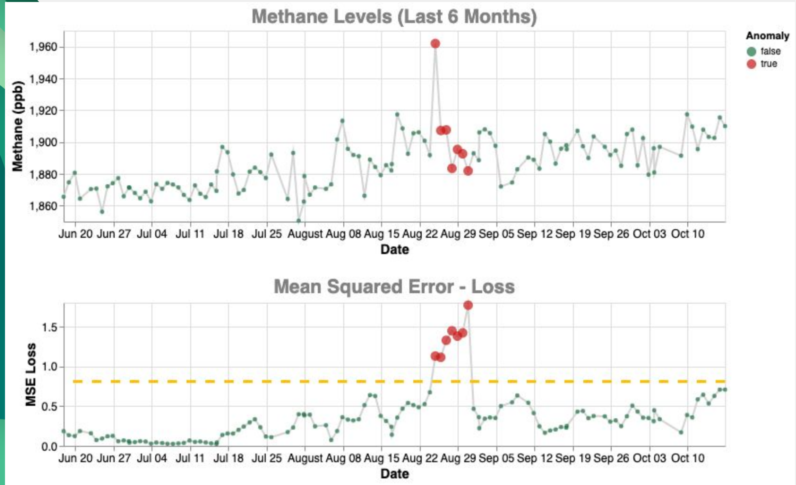


Appendix - Chico - Synthetic Test - Single Anomaly

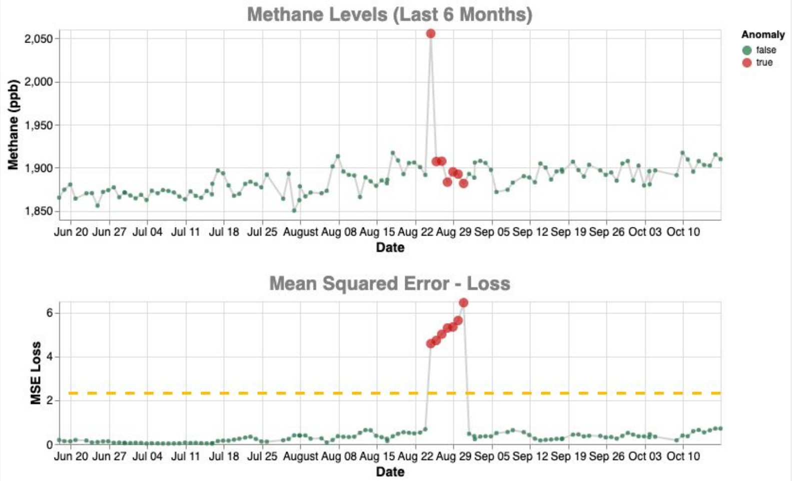
True



5% + mean



10% + mean



Appendix - Abbreviations

EPA - Environmental Protection Agency

IPCC - Intergovernmental Panel on Climate Change

Appendix - Product Architecture