TRANSFORMERS MAGAZINE'S
INDUSTRY NAVIGATOR

INVESTMENTS, ARTIFICIAL INTELLIGENCE AND SUSTAINABILITY

CONFERENCE 2024

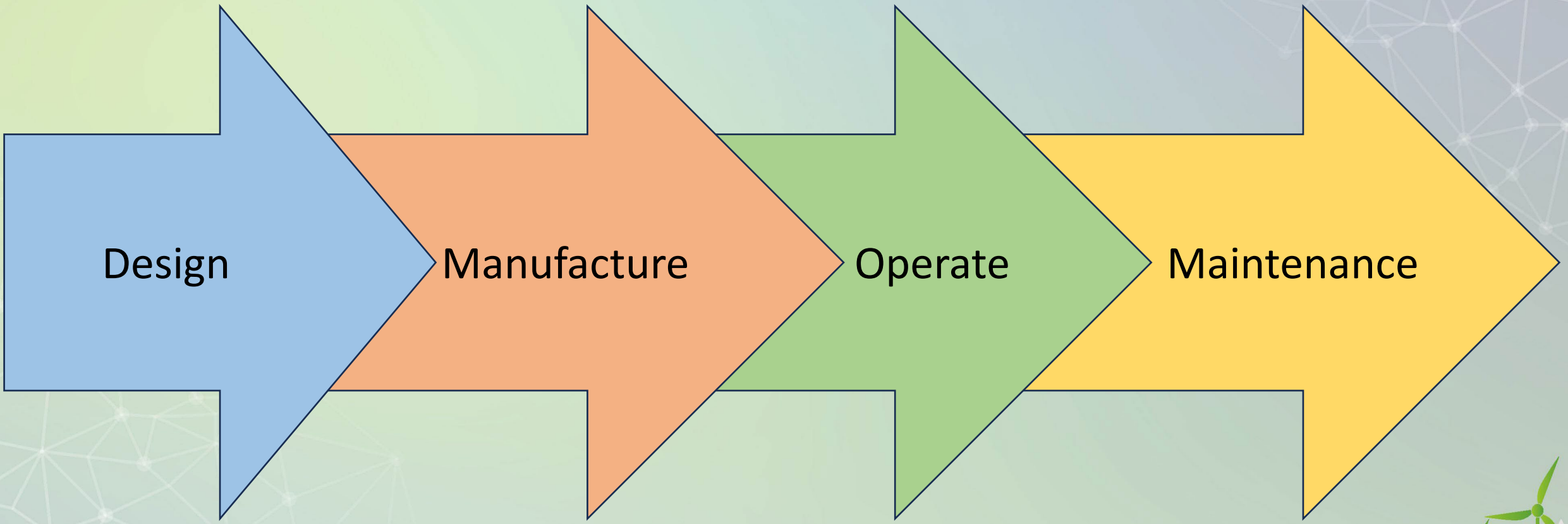
Synthetic Dissolved Gas Analysis Data Generation and Training of a Time Series Predictor

Mauricio Soto, PhD

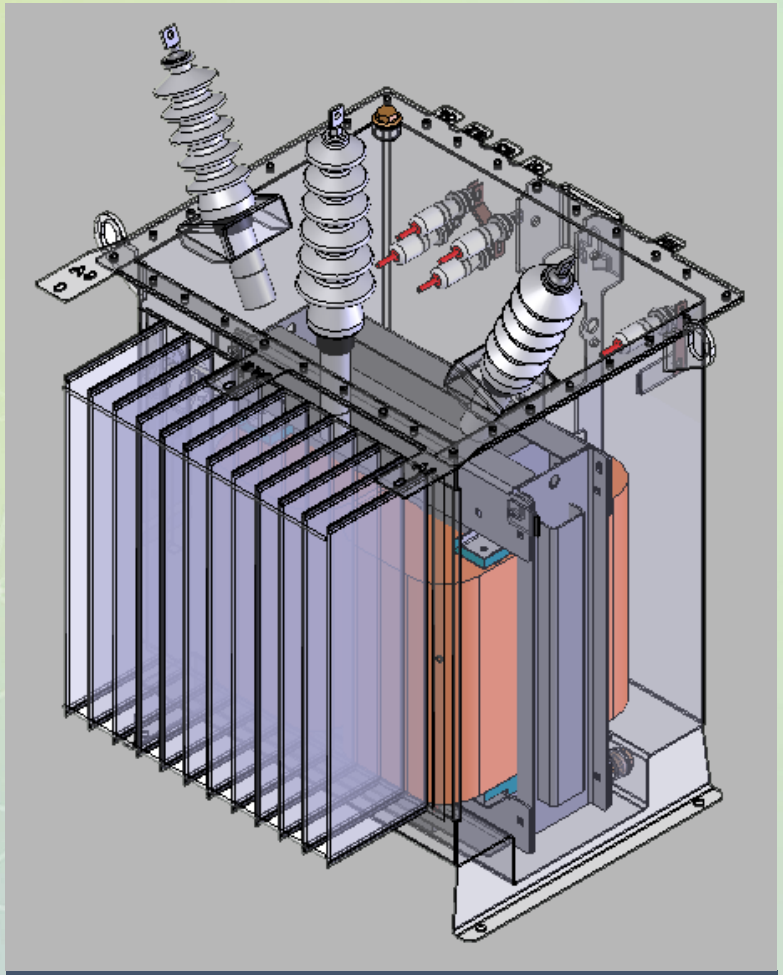
June 10th 2024

AI can increase accuracy, productivity, reliability and safety throughout the life cycle of transformers

AI Can Add Value Throughout the Life Cycle of the Transformer



Design



Electrical and Mechanical Design Optimization

Narsinha Engineering Works
Manufacturer & Repairers of Power & Distribution Transformers

Ref. NEWN/QTN./**024**/2013-14. Dt: 07 Nov 2013

To,
ELECTRALLOY SPECIAL STEEL CASTINGS PVT LTD.
Hosur.

Kind attn: MOHAN N Sir.
Sub. : **Quotation for supply of 500 KVA, 11/ 0.433 kV Transformer.**

Dear sir,

With reference to above, we are pleased to quote our most competitive rates for supply Of above mentioned transformer suitable for outdoor duty. As per IS 2026. With following Specification.

KVA	**500 Copper Wound**
HV Voltage	11000
LV Voltage	433
Frequency	50 Hz.
Phases	3
Connections	HV – Delta, LV – Star
Tappings	On HV side by means of off load Tap Changer, +_5%
Vector Group Ref.	Dyn11
No load losses	As per IS
Load loss at 75 deg. C	As Per IS
% Impedance at 75 deg. C	5 %
Temp. Rise oil / winding	45 deg. C / 50 deg. C
Type of cooling	ONAN
Fittings & accessories	All standard accessories as per IS 2026
Price	Rs.**5,30,000** (Five lakhs Thirty Thousand only)
Taxes	**As per applicable**
Transportation	**In Your Scope.**
Guarantee	One YEAR From the date of Commissioning
Delivery	03 to 04 Weeks from the receipt confirmed order along with advance.
Terms of payment	**50%** advance along with order balance against proforma invoice but prior to dispatch.

Hope we have submitted sufficient information for your scrutiny, should you require any further clarification / information please call on us and same will be submitted immediately.
Thanking you,

For Narsinha Engineering Works,

Partner

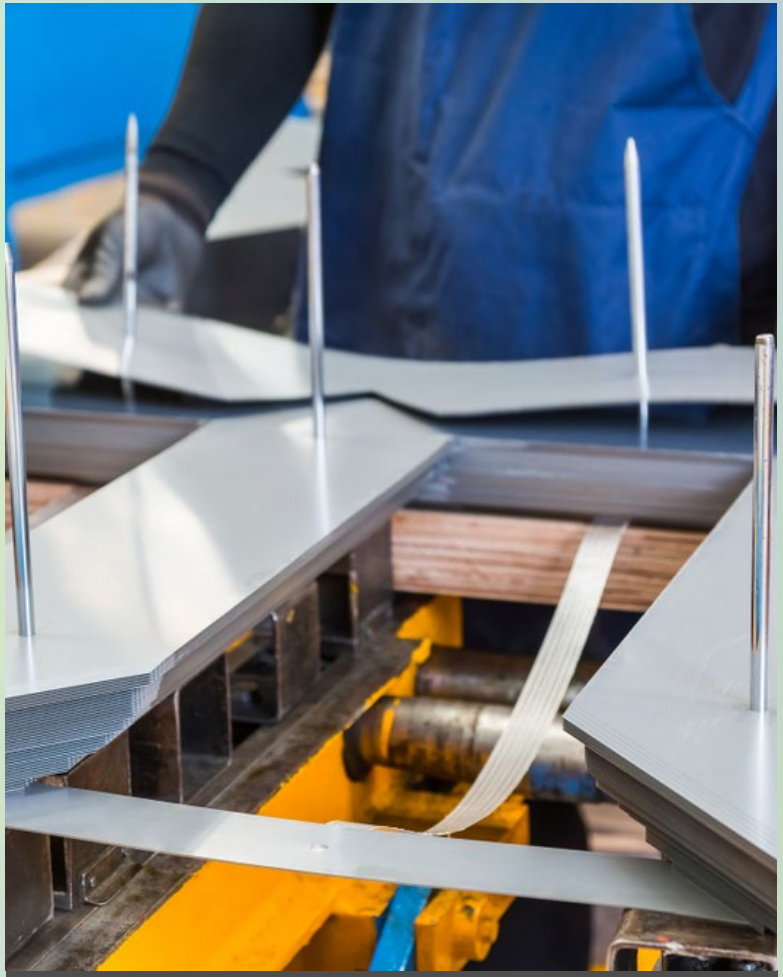
Address: Plot No. E-35, M.I.D.C Area, Nanded (MH)-431603
Contact: 02462-286529
Mo: 09960190080/09527248498
E-mail: narsinhaenggworks@gmail.com
Website: www.narsinhatransformers.com

Transformer Specification Support

Failure Prevention

Manufacture

Process Optimization



Manufacturing Safety

Assisted Inspection

Operate

Failure Detection

Behavior Forecasting

Increased Security

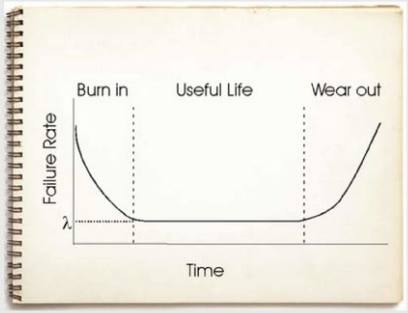
Maintenance

Internal Causes

- Overloading
- Lightning Surge
- Deterioration of Insulation
- Improper Cooling
- Moisture

External Causes

- Extreme Weather
- Line Surges
- Poor Workmanship
- Inadequate Maintenance
- High-Voltage Disturbance
- Malicious Mischief



Failure Clustering

Predictive Maintenance

Chat Bot Support

All These AI Tools Have One Thing in Common:

They All Need Data

(and usually, lots of it)

The Quality of AI Models is Proportional to the Data They are Trained On

In the Transformers Community, Due to Privacy, Security, Compliance and Competitiveness Concerns, Data is Not Freely Shared

Data Synthesis

GANs
Regression
Bayesian models
Markov chains
Time series
Etc.

Data Anonymization

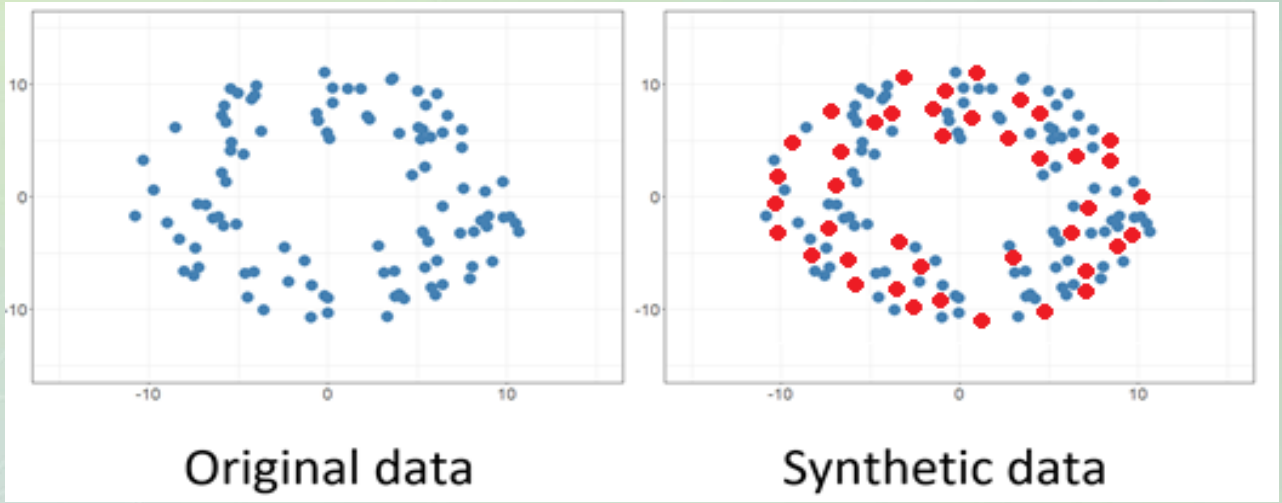
Synthetic Data for Time Series

Synthetic data is artificially generated rather than produced by direct measurement

When we don't have access to real-world data, we create data that behaves how a transformer would behave

We use a combination of real transformer data, domain knowledge expertise, and statistical tools

Use cases: Train ML models, perform testing of monitoring algorithms, perform data anonymization, transformer simulation, and other digital functionality (data modeling, data analysis, etc.)



Our Approach (Synthetic Data Generation for Time Series)

Generating synthetic data for time series

- Time Series: Datapoints indexed in order of time
- Dissolved gas analysis, load, temperature, tap position, vibroacoustic, frequency response analysis, etc.

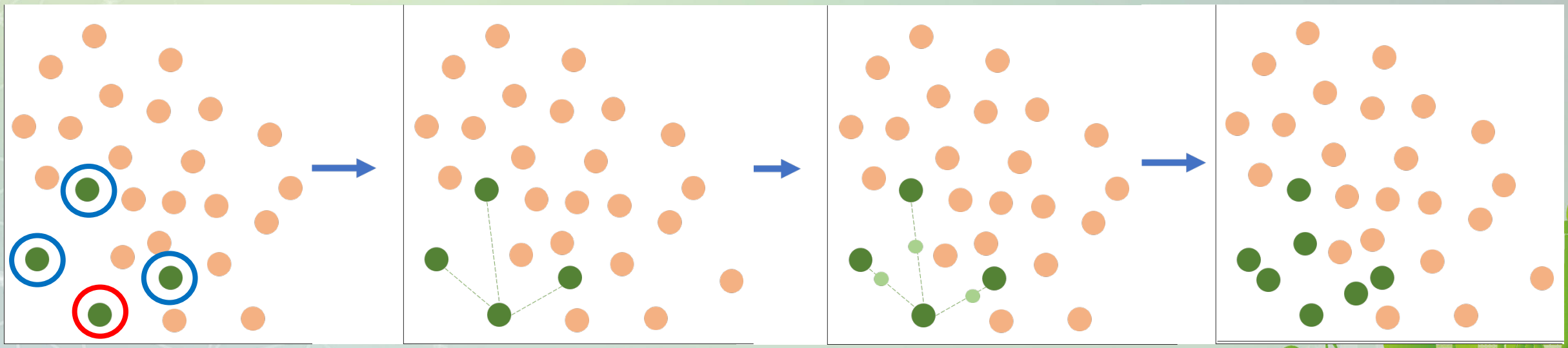
Input: A dataset to use as a sample of behavior

Output: A different dataset that follows a similar behavior as the input dataset

SMOTE (Synthetic Minority Oversampling Technique)

Used in classification problems

- For each datapoint in the minority class:
  - Select k nearest neighbors of the minority class
  - Calculate a vector between the datapoint and its neighbors
  - Multiply each vector by random number to change magnitude
  - Create new datapoint with the resulting magnitude



Our Approach

Adapted SMOTE to be applied to time series:

- Flatten search space to a single dimension (time)

Added an additional multiplier to increase flexibility

Added domain knowledge filters to ensure the data behaves as a real transformer would

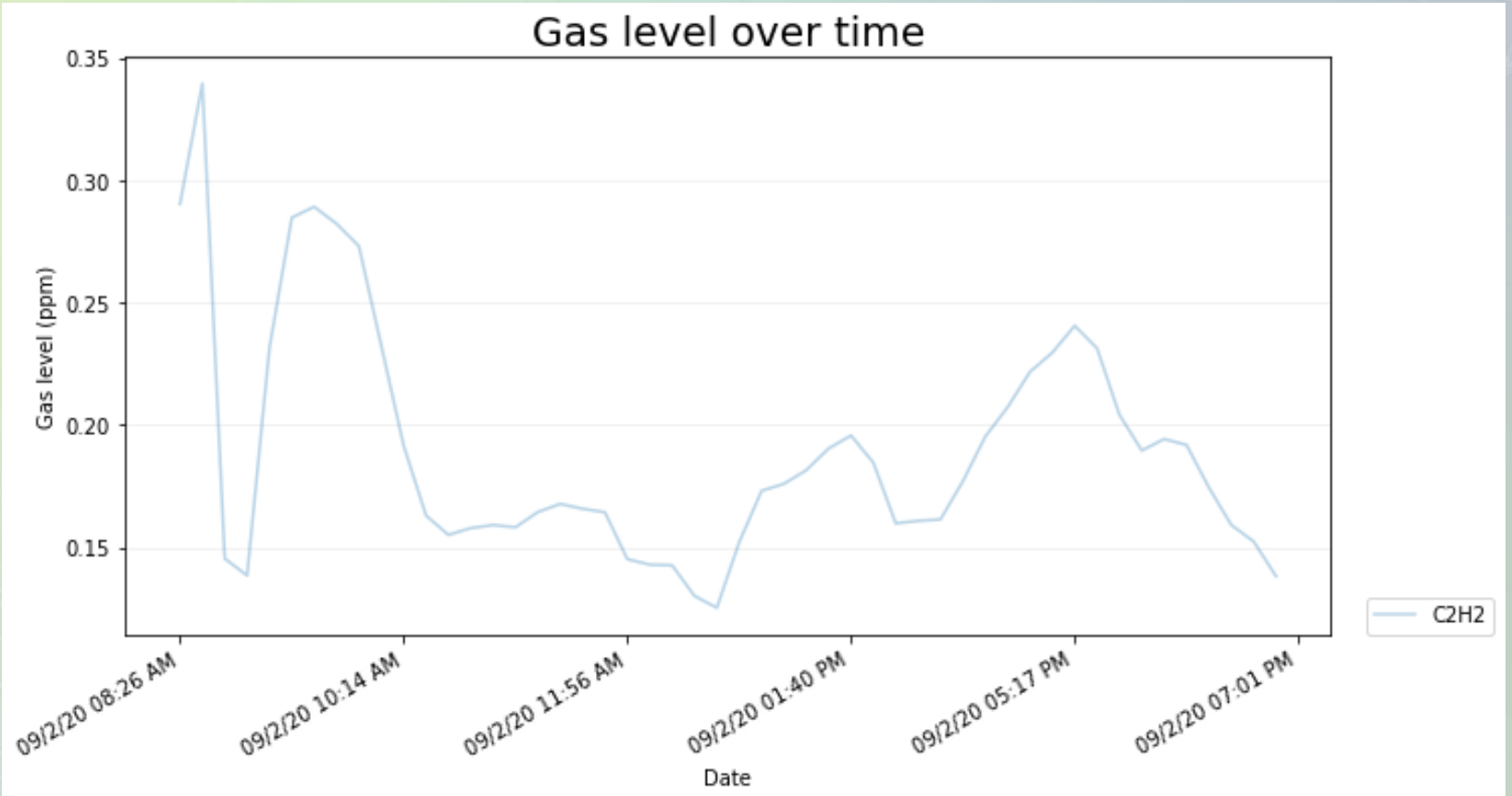
- E.g., DGA values can't be negative; load has periodicity per day, week and year; lower temperature increases oil viscosity; ambient temperature vary by location and time of year; etc.

Data Synthesis Procedure Pseudocode

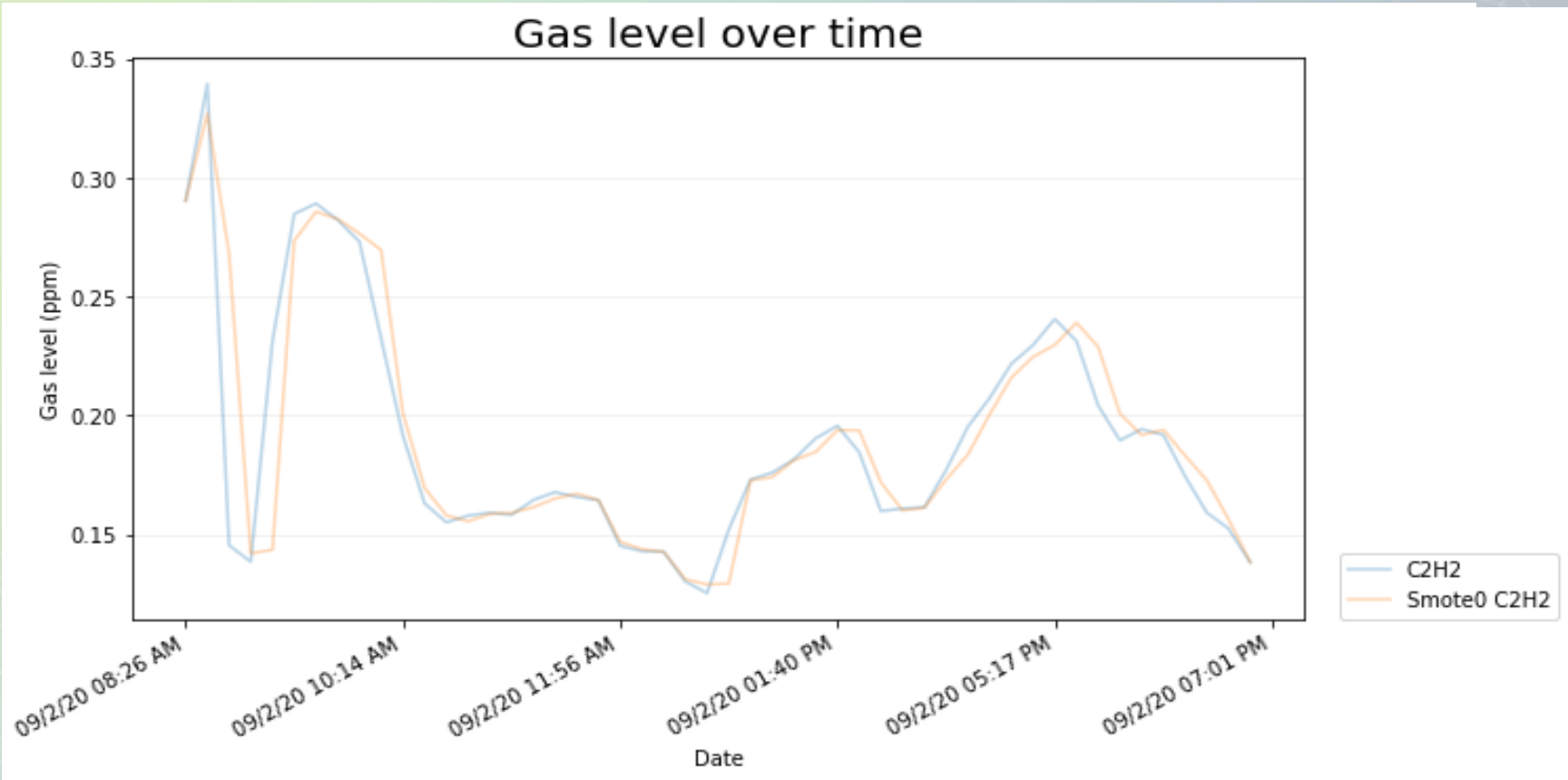
```
1.  mult = 1 # fixed multiplier
2.  procedure Synthetic_Data_Generation():
3.      time_series = get_original_dataset()
4.      new_ts = [] # new synthetic time series list
5.      for each p in time_series:
6.          diff = (p.next - p)
7.          rand = random_number() # range [0,1]
8.          new_p = p + (diff * rand * mult)
9.          new_p = apply_filters(new_p)
10.         new_ts.add(new_p)
11.     return new_ts
```

Algorithm 1. Proposed synthetic data generation methodology

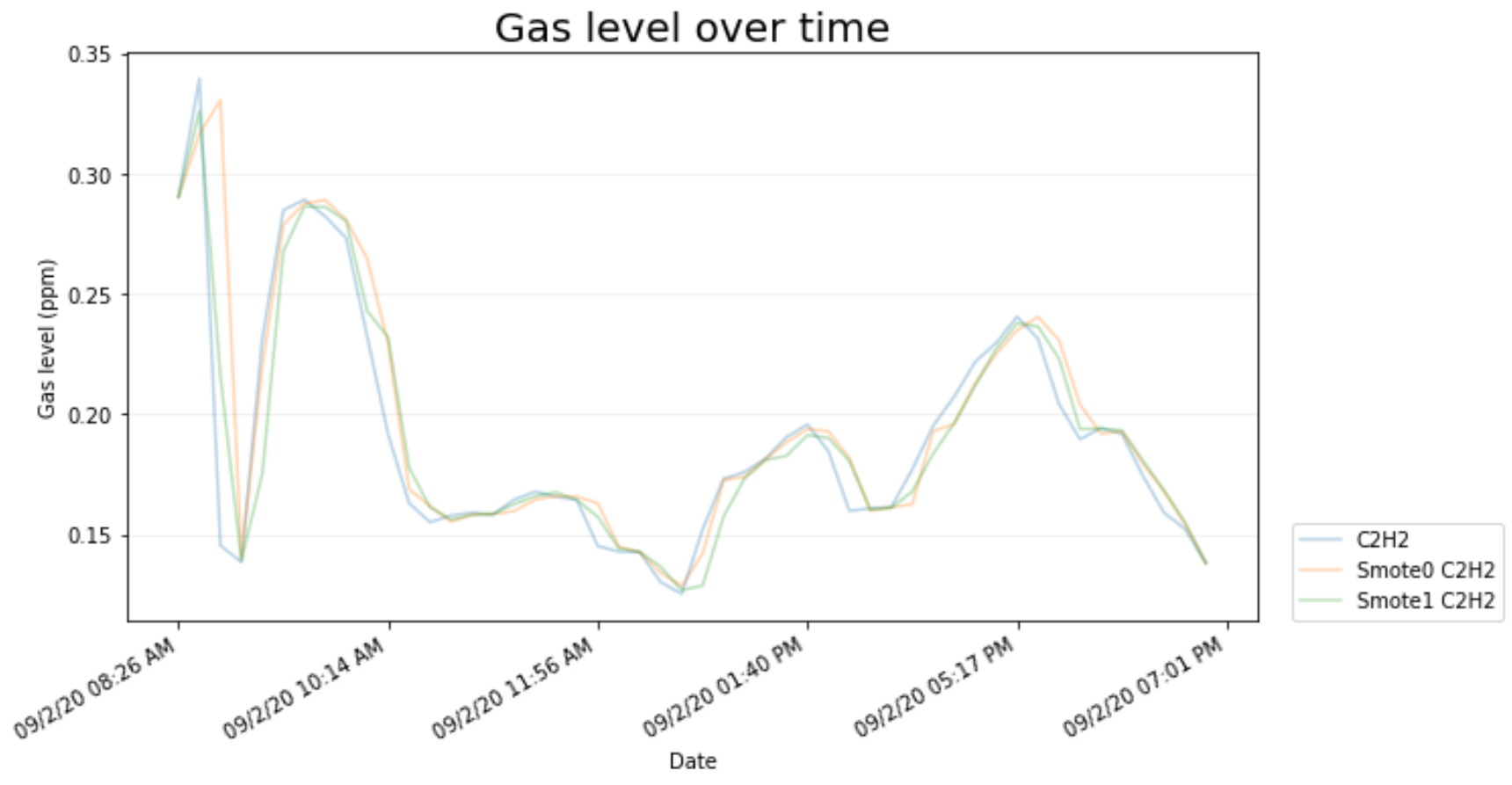
Synthesis Technique



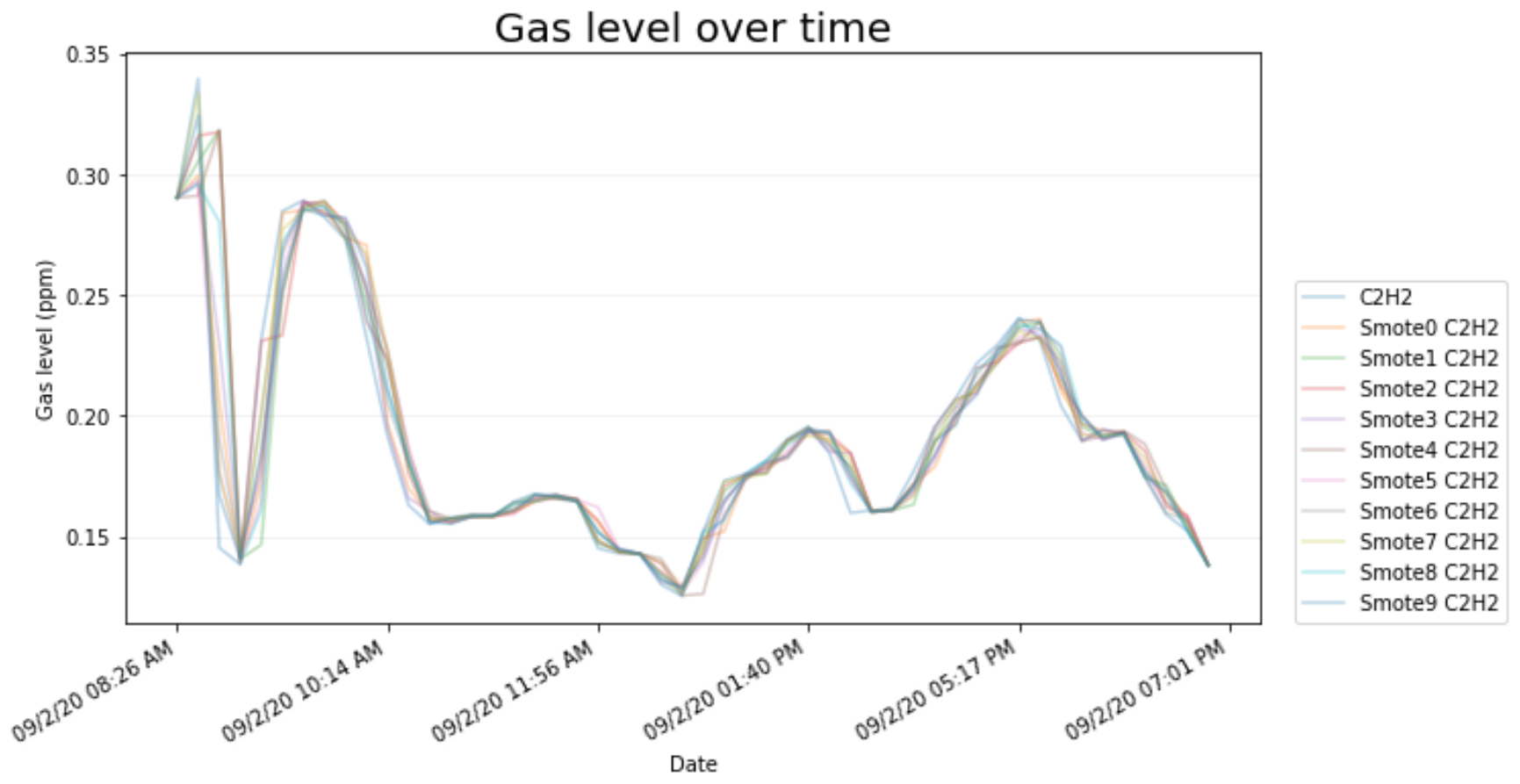
Synthesis Technique – One Synthetic Time Series



Synthesis Technique – Two Synthetic Time Series

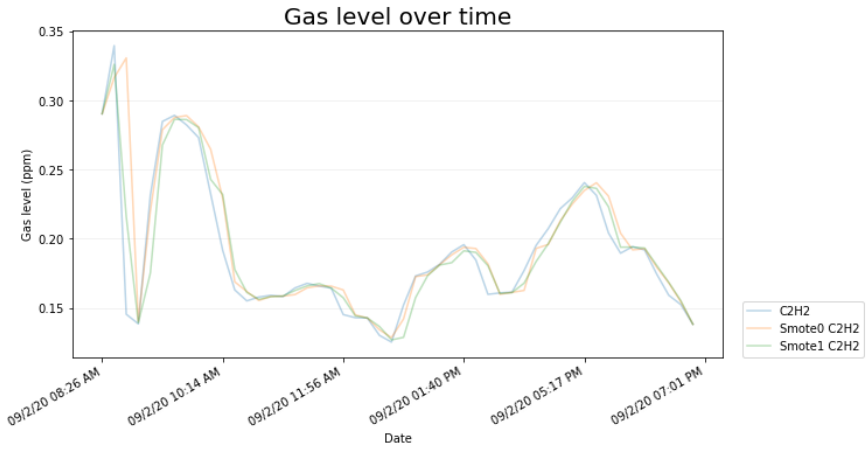


Synthesis Technique – Ten Synthetic Time Series

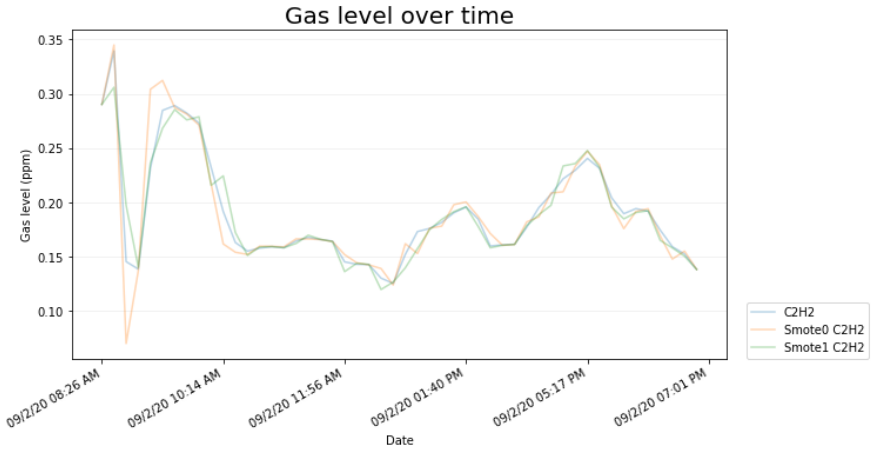


Varying Configurable Multiplier

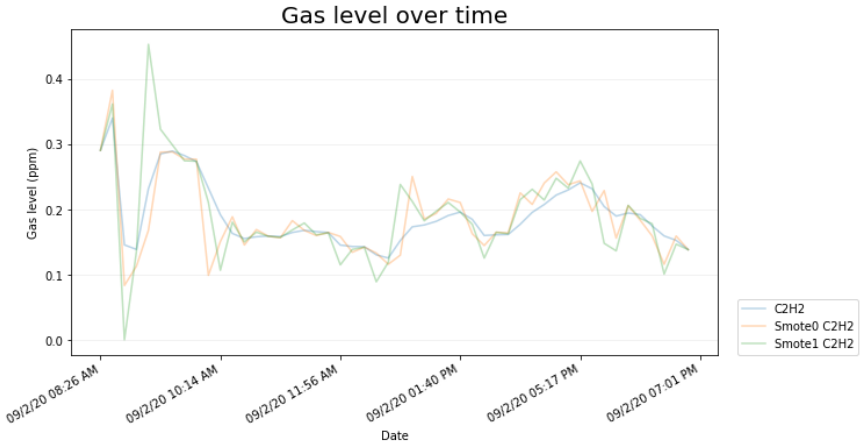
Multiplier = 1



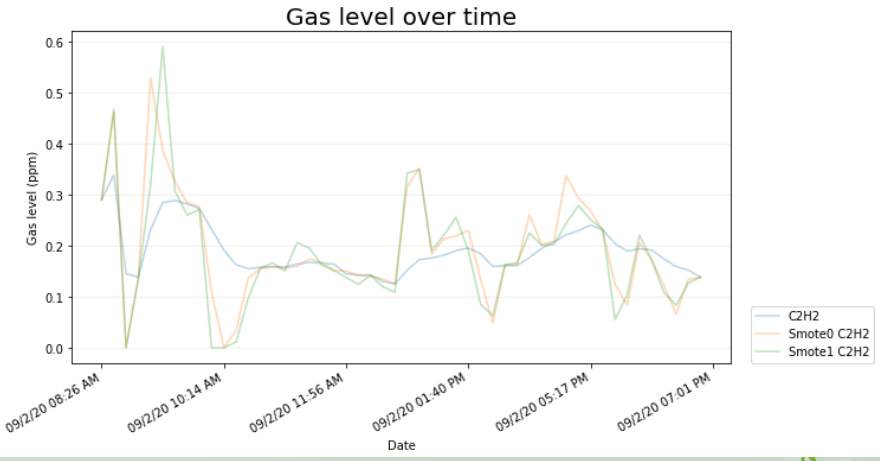
Multiplier = 2



Multiplier = 5



Multiplier = 10



Exploring Uses of Synthetic Data: Predicting DGA Readings Using a Synthetic Corpus

Sample Transformer Data

- Three Transformers
- M10 sensor / Samples every 10 minutes
- Four months and a half, for each transformer
- ~18,500 readings per transformer
- Used C57.104 to divide between typical and atypical readings

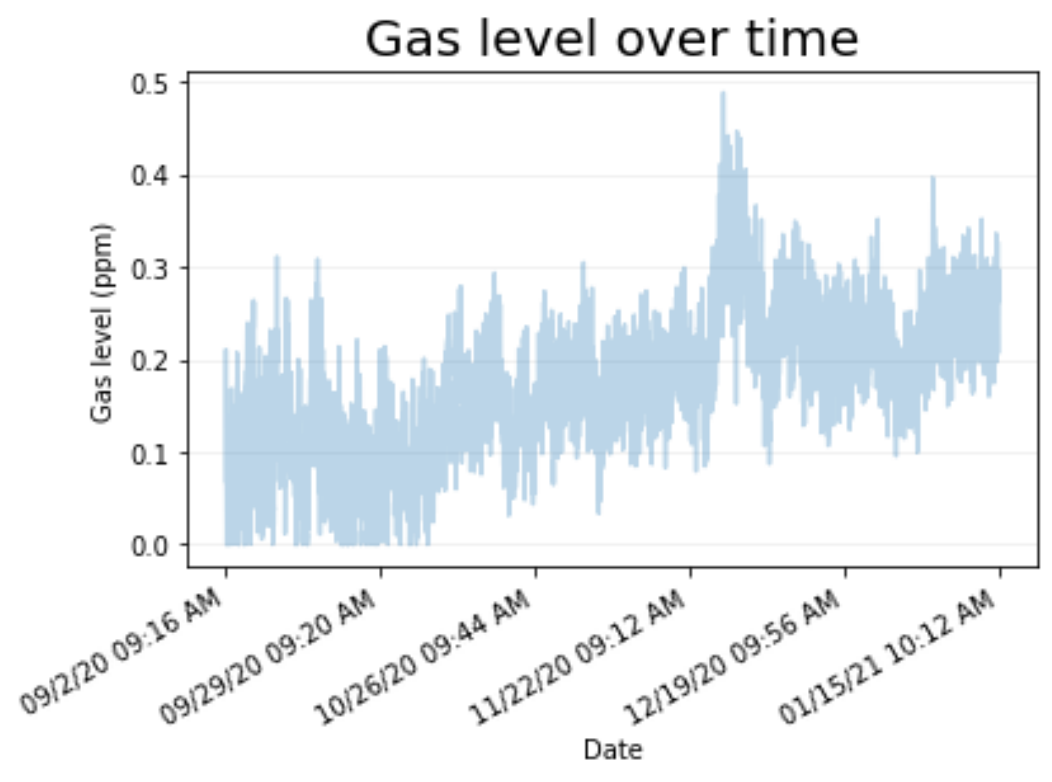
Table 1—90th percentile gas concentrations as a function of O_2/N_2 ratio and age in µL/L (ppm)

		O_2/N_2 Ratio ≤ 0.2				O_2/N_2 Ratio > 0.2			
		Transformer Age in Years				Transformer Age in Years			
		Unknown	1 – 9	10 – 30	>30	Unknown	1 – 9	10 – 30	>30
Gas	Hydrogen (H_2)	80	75		100	40	40		
	Methane (CH_4)	90	45	90	110	20	20		
	Ethane (C_2H_6)	90	30	90	150	15	15		
	Ethylene (C_2H_4)	50	20	50	90	50	25	60	
	Acetylene (C_2H_2)	1	1			2	2		
	Carbon monoxide (CO)	900	900			500	500		
	Carbon dioxide (CO_2)	9000	5000	10000		5000	3500	5500	

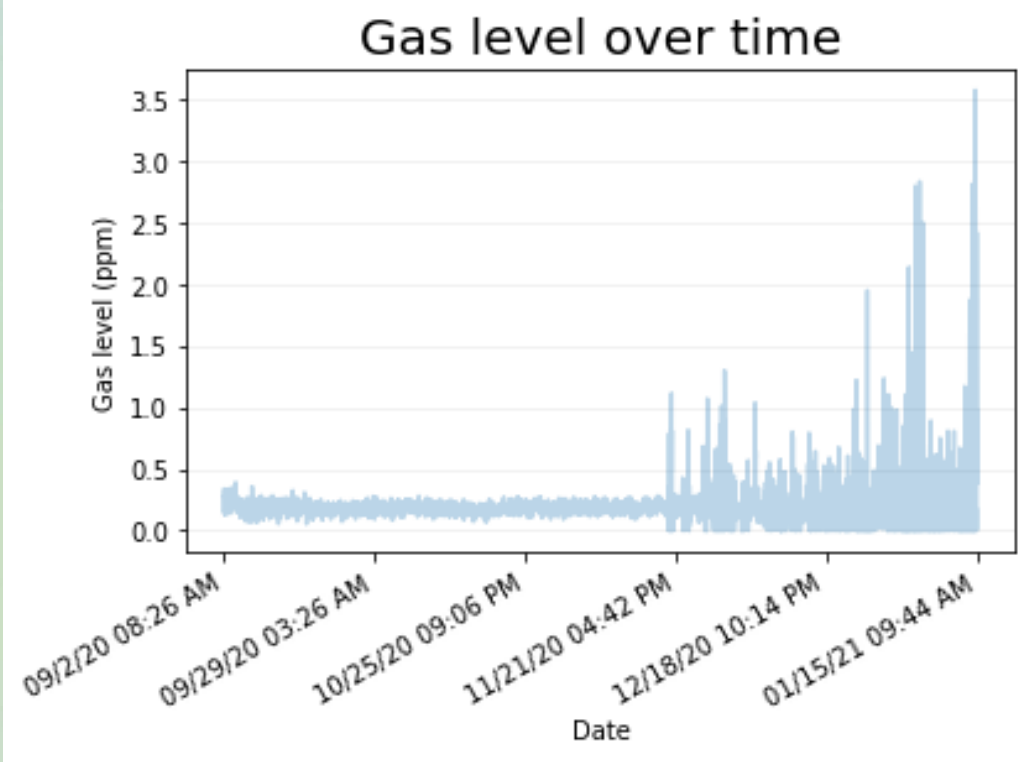
NOTE—During the data analysis, it was determined that voltage class, MVA, and volume of mineral oil in the unit did not contribute in significant way to the determination of values provided in Table 1.

Sample Transformer Data (Acetylene)

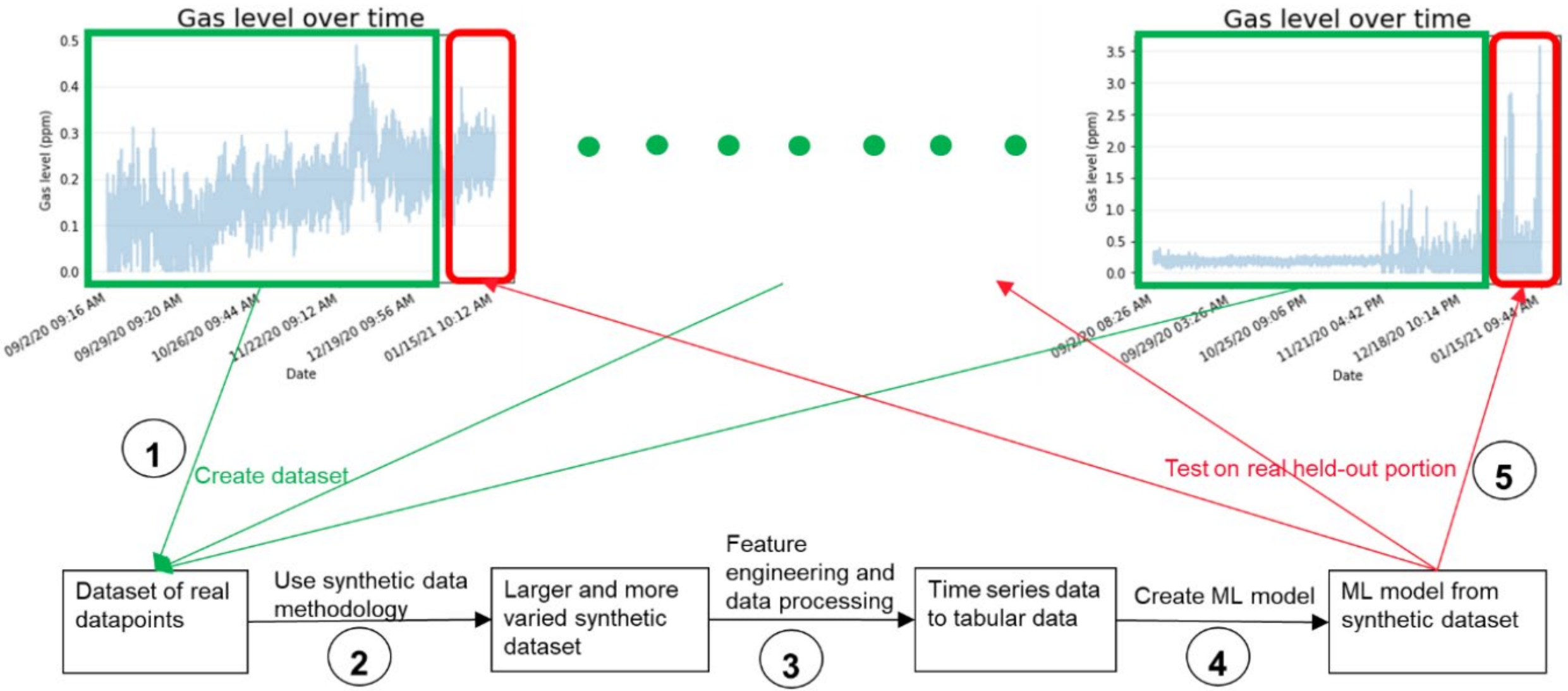
Typical time series



Atypical time series



Split Into Training and Testing Sets (90/10)



Test Classification Model on Held-Out Testing Data

Accuracy: 98.03%

Balanced accuracy: 93.97%

- Typical Readings: **Precision**: 99.75% **Recall**: 98.22%
- Atypical Readings: **Precision**: 54.24% **Recall**: 89.72%

Atypical readings:

Recall:

- From the readings that were actually atypical, how many did our model said were atypical
- 89.72% of atypical readings where correctly reported as atypical

Precision

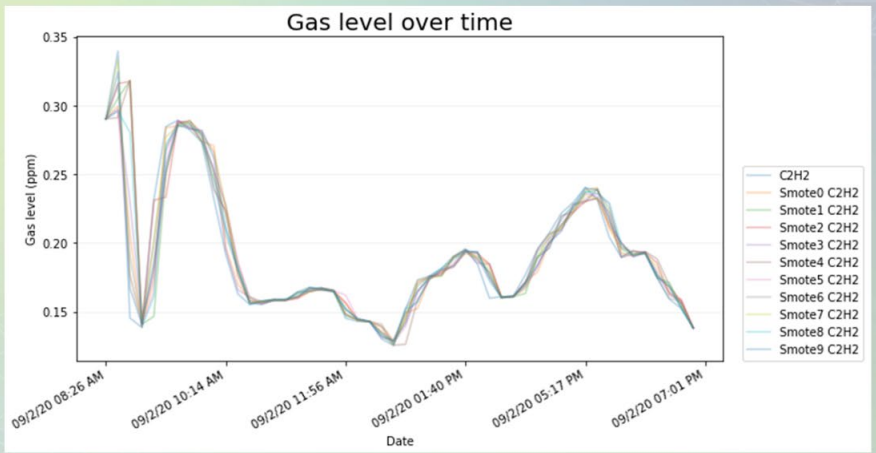
- From the readings our model said where atypical, how many where actually atypical
- Reported a lot of false positives: 45.76%

Take Aways

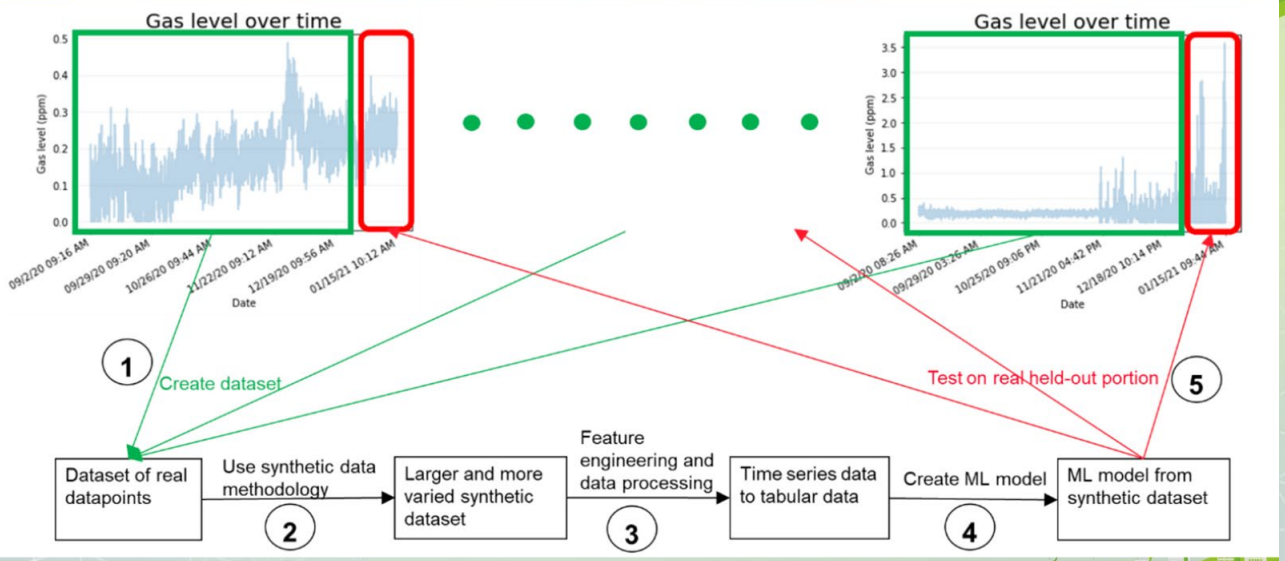
- Proposed a methodology to generate transformer synthetic data using a combination of real transformer data, domain knowledge expertise, and statistical tools
- Methodology can be used to any time series such as DGA, load, temperature, tap change positions, vibroacoustic, etc.
- Use cases: ML modeling, perform testing of monitoring algorithms, data analysis, data anonymization, transformer simulation, etc.

In the Transformers Community,
Due to Privacy, Security, Compliance
and Competitiveness Concerns,
Data is Not Freely Shared

Synthesis Technique – Ten Synthetic Time Series



Split Into Training and Testing Sets (90/10)



Take Aways

- Proposed a methodology to generate transformer synthetic data using a combination of real transformer data, domain knowledge expertise, and statistical tools
- Methodology can be used to any time series such as DGA, load, temperature, tap change positions, vibroacoustic, etc.
- Use cases: ML modeling, perform testing of monitoring algorithms, data analysis, data anonymization, transformer simulation, etc.