Data Augmentation with Synthetic Data for Enhanced Time Series Forecasting

Paul BERHAUT - Research Engineer - Air Liquide R&D
paul.berhaut@airliquide.com

Air Liquide : A world leader in gases, technologies, and services for…

INDUSTRY

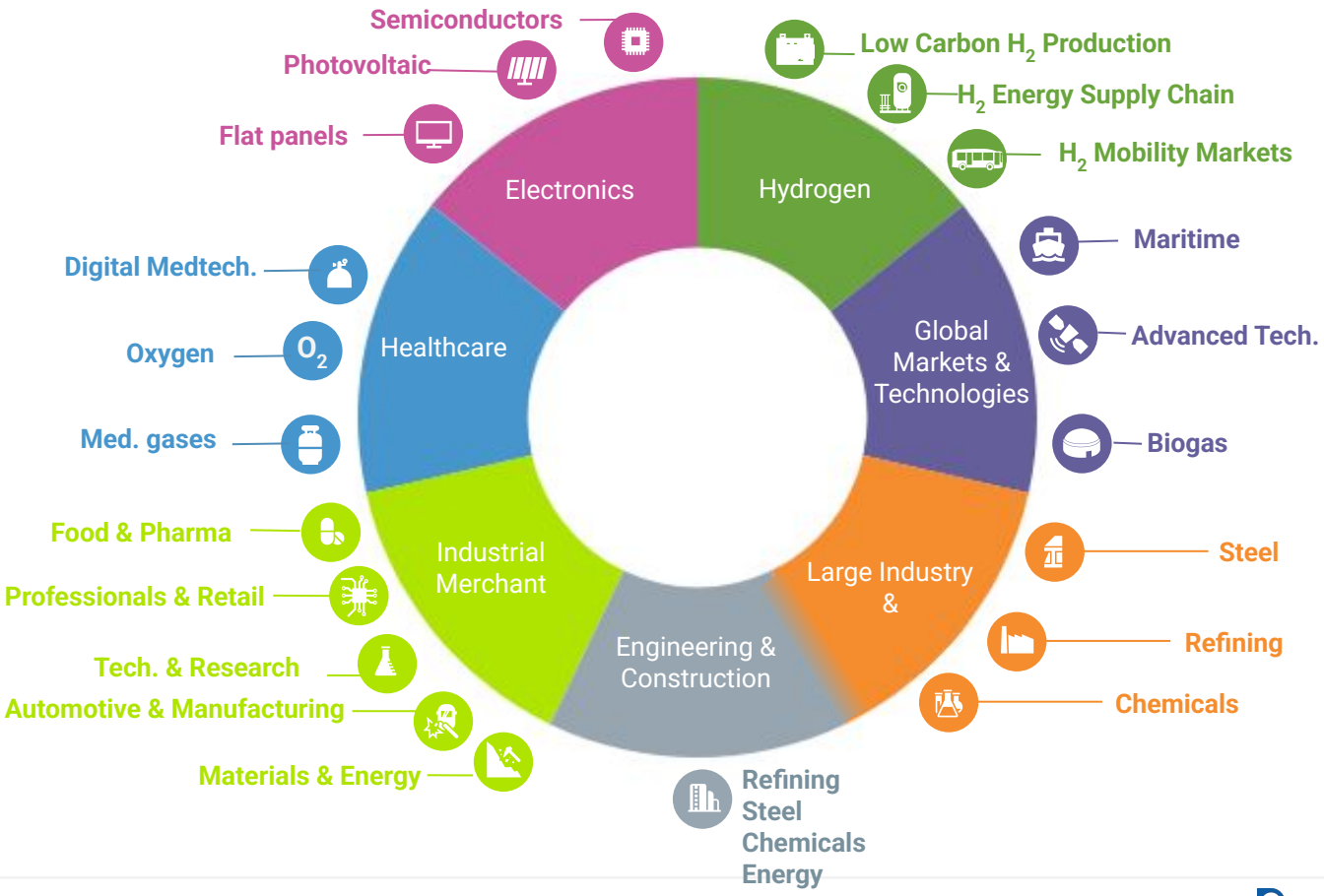
Sustainable solutions for a wide range of industrial processes for customers (energy, metals, food, chemicals, automotive, pharmaceuticals, etc.) **and for transportation**

HEALTH

Patients at home
Hospitals
Specialty ingredients

Air Liquide R&D - Innovation across diverse activities

from
PRODUCTION, DISTRIBUTION
to
APPLICATIONS



Data science @ Air Liquide

Data Sources

Industrial Assets

Logistics

Patients & Customers

Infrastructure

Core Data Models
Data Platforms
AI Platforms

Data Science & AI Use-Cases

Predictive Maintenance

Supply Chain Optimization

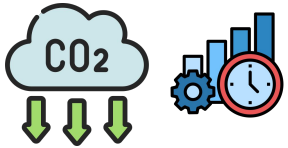
Energy Management

Digital Marketing

Objectives

Patient & Customer Satisfaction

Industrial Performance



Data Augmentation for Time Series

- **Data augmentation is often used when applying machine learning models to images.**
  - Geometric transformations: scaling, cropping, flipping, translating, …

- **Augmentation is expected to improve a model's generalization capability**

- **But Time Series data requires preservation of sensitive features (autocorrelation, trends, seasonality, …)**
  - Will the advantage of **more data** outweigh the negative of **lower quality data** ?

(Gao et al, https://arxiv.org/abs/2310.10060)
(Fawaz et al, https://arxiv.org/abs/1808.02455)
(Iwana et al, https://arxiv.org/abs/2007.15951)

Goal of the project

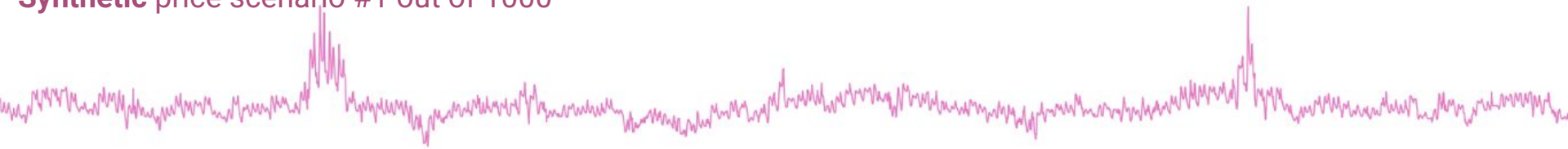
- **Air Liquide has an internal tool to generate synthetic scenarios.**

- **We have used this tool on open data: electricity prices in France from 2016 to 2020.**
  - → 616 years of synthetic data available

- **We would like to estimate the applicability of this tool as a data augmentation tool.**
  - Is it better than *"no augmentation"* ?
  - Is it better than *"basic augmentation techniques"* ?
  - *"basic augmentation techniques"* to be proposed by students as part of the project.
  - Optional: Test and/or propose other advanced augmentation techniques (TimeGAN or other?)
  - Proposed task to evaluate models: forecast electricity prices 6h, 12h, 24h, 48h and 72h and 168h into the future.

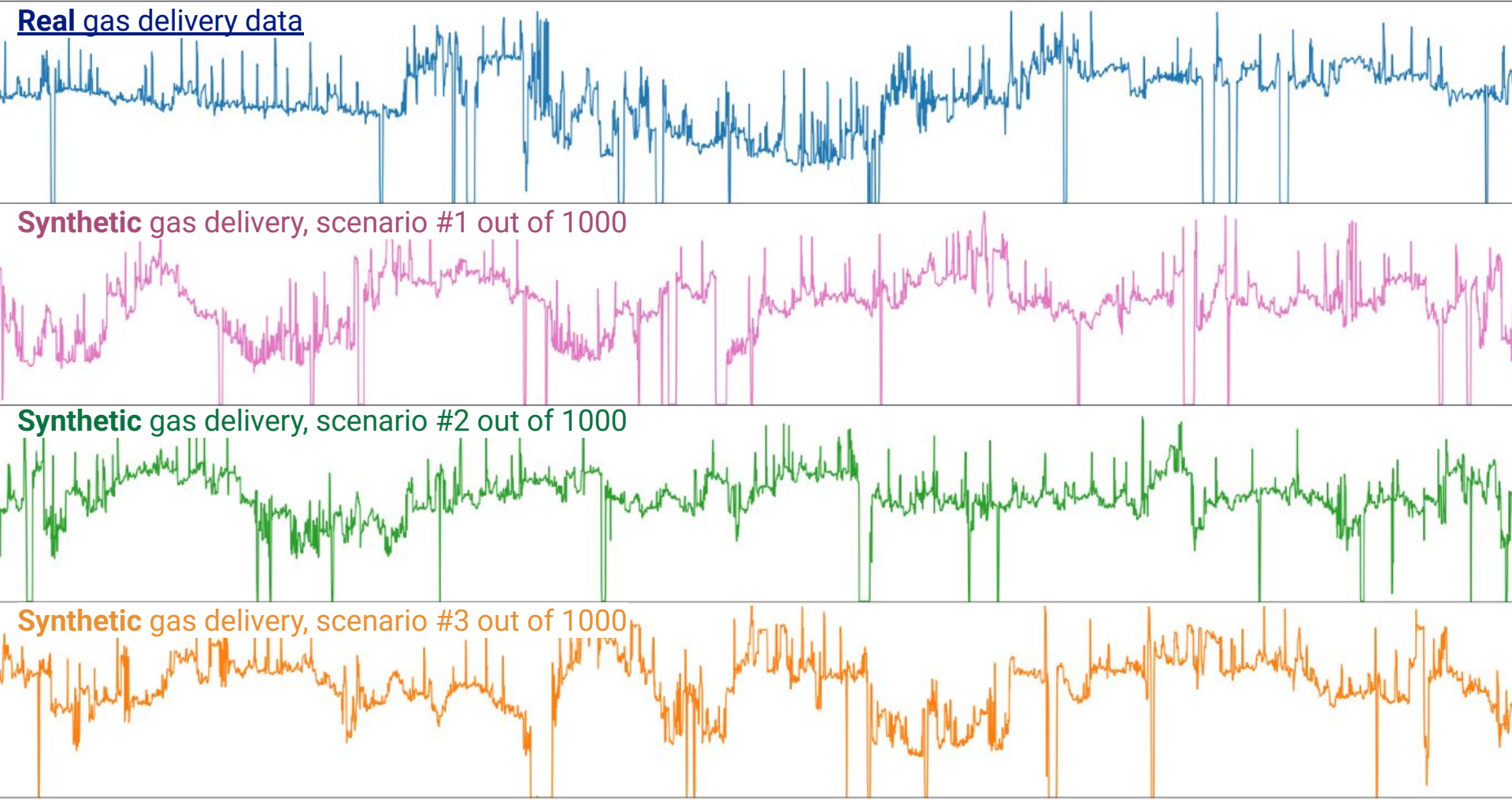
Real historical electricity price data

Synthetic price scenario #1 out of 1000

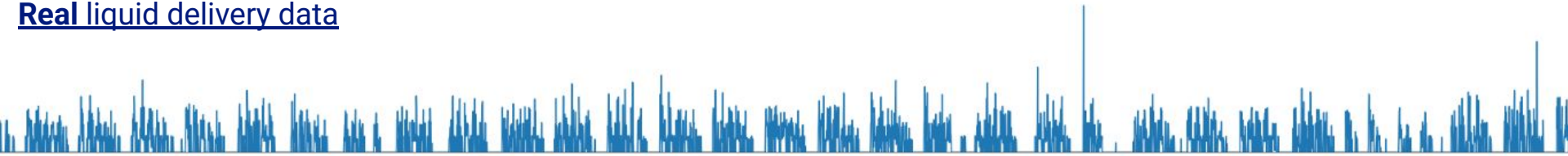


Synthetic price scenario #2 out of 1000

Synthetic price scenario #3 out of 1000

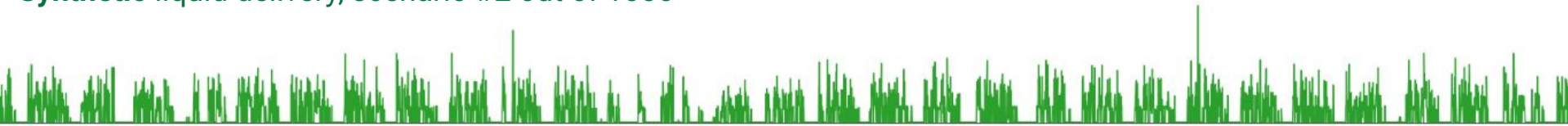

Real gas delivery data
Synthetic gas delivery, scenario #1 out of 1000
Synthetic gas delivery, scenario #2 out of 1000
Synthetic gas delivery, scenario #3 out of 1000

Real liquid delivery data



Synthetic liquid delivery, scenario #1 out of 1000

Synthetic liquid delivery, scenario #2 out of 1000



Synthetic liquid delivery, scenario #3 out of 1000