Air Liquide
creative oxygen

Uncertainty Quantification for Customers Demand Forecasting

SafeAI Workshop - AAAI 2022

SystemX INSTITUT DE RECHERCHE TECHNOLOGIQUE

IRT SAINT EXUPÉRY

Confiance ai

THIS DOCUMENT IS ••CONFIDENTIAL-RESTRICTED

Marc Nabhan (Air Liquide R&D), Kevin Pasini (IRT SystemX), Luca Mossina (IRT Saint Exupéry), Mouhcine Mendil (IRT Saint Exupéry)

DISTRIBUTION LIST : SafeAI Workshop - AAAI 2022

AAAI-22

SafeAI

Air Liquide Demand Forecasting Use Case

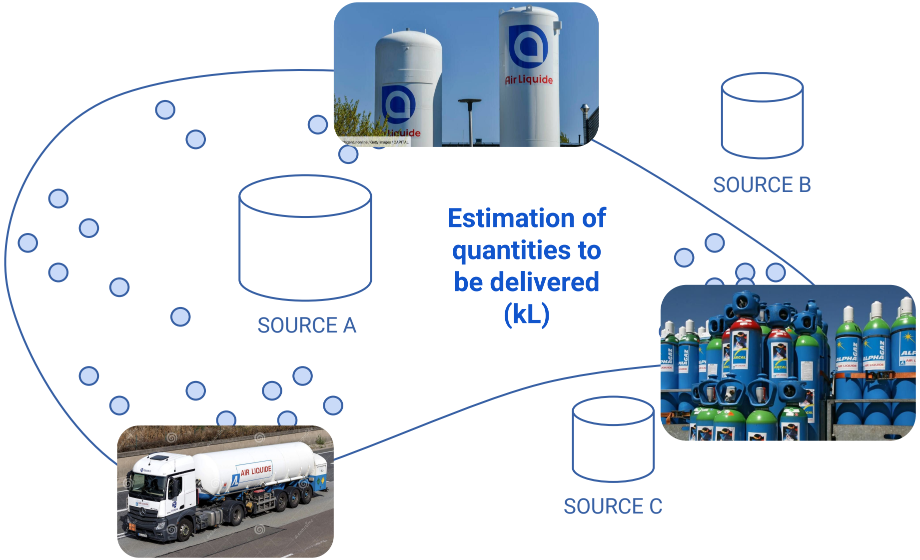
CONTEXT
Predicting the future trends of customer demands, **as precisely as possible**, is highly useful for the production sites

DATA - INFORMATION - KNOWLEDGE
Multivariate **time series**: historical consumption, geographical distribution, customers, orders, contextual data, seasonality

	Sourcing Decision	Product Loaded	Sourcing Deviation
SOURCE A	kL	kL	+%
SOURCE B	kL	kL	-%
SOURCE C	kL	kL	+%
Total	kL	kL	0%

Deviation from Sourcing Plan

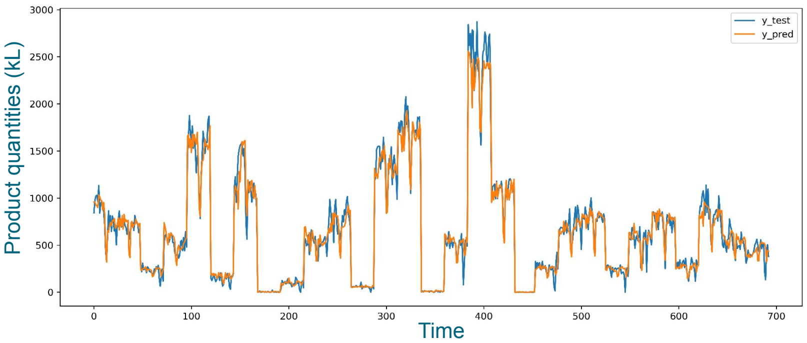
Perfect prediction



Air Liquide Demand Forecasting Use Case

INFERENCE ENGINE (SOLUTION V0)
XGBoost Regression - Single point predictions

OBJECTIVE
Prediction Intervals that quantify uncertainty in the forecast with **high guarantee**



Technical estimation	**300 kL**
Forecast estimation	**270 kL**

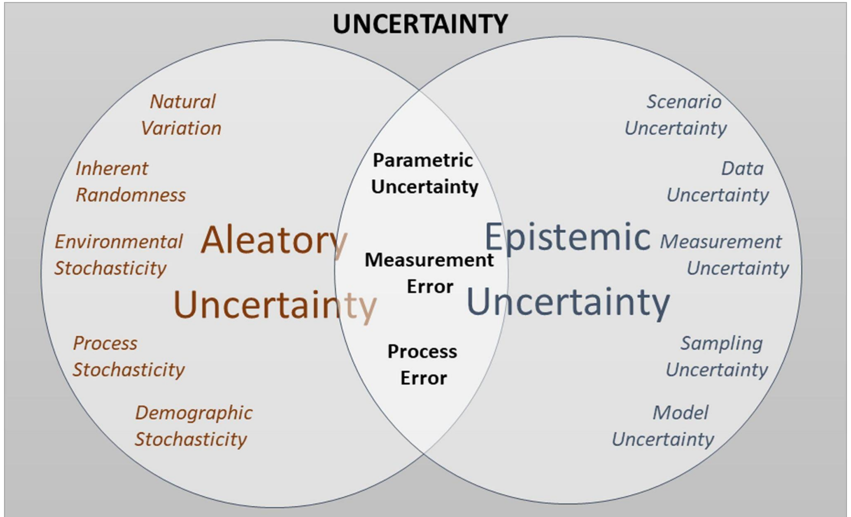
Technical estimation	300 kL
Minimum guaranteed	**220 kL**
Forecast estimation	270 kL
Maximum guaranteed	**320 kL**

Uncertainty: Definition

Uncertainty in the Machine Learning process can be reduced to two types [Hüllermeier and Waegeman, 2021]:

- **Aleatoric uncertainty** (*aka* statistical uncertainty), which is ***irreducible*** as due to the **intrinsic randomness** to the phenomenon being modeled. (Coin flipping)

- **Epistemic uncertainty** (*aka* systematic uncertainty), which is ***reducible*** because it is caused by a **lack of knowledge** of the best model, **insufficient information** or **lack of data**. (No seasonality data)



[Shamsi et al., 2020]

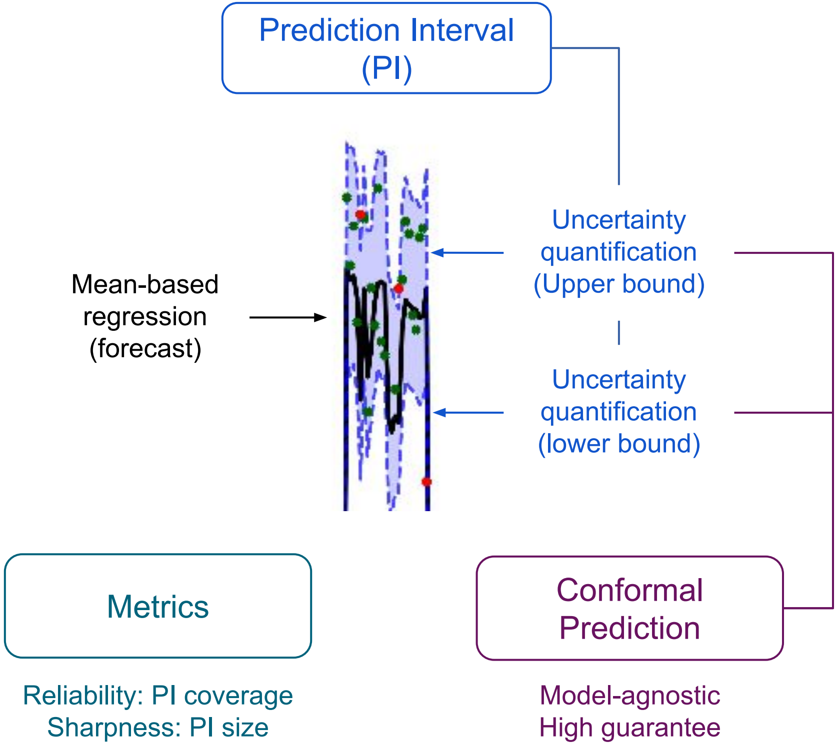
→ It is usually **not that simple to distinguish** the 2 sources of uncertainty in Machine Learning.

Confiance.ai work package

1. **Uncertainty Quantification (UQ)**
   - Quantile Regression
   - Variance Estimation

2. **Evaluation & Metrics**
   - Reliability: PI Coverage
   - Sharpness: PI Size

3. **Conformal Prediction (CP)**
   - Mean-based vs Quantile-based
   - Data Scheme



Quantile Regression: Definition

- Standard way to build prediction intervals

- Aims at **predicting the conditional quantiles** of the response variable (in contrast to the single-value regression that estimates the conditional mean)

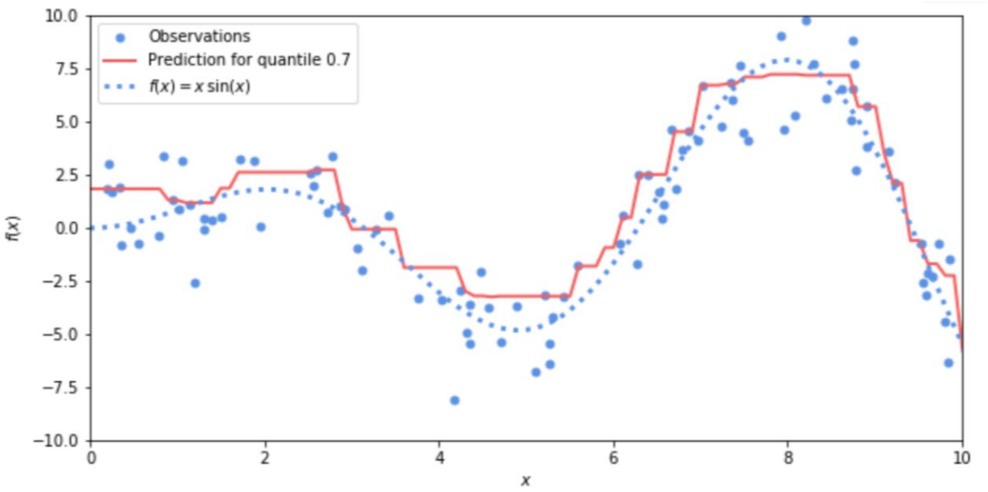
- **Quantile definition:** Given a random variable **Y** and a value **p** in [0 , 1], the associated quantile **q** is the value such that:

$$P(\mathbf{Y} \leq \mathbf{q}) = \mathbf{p}$$

  → The median is the 50th quantile

- **Boosting algorithms** (e.g. Gradient Boosting) are examples of regressors that can be transformed into **quantile regressors** with the **quantile loss**

Quantile Regression: Example



The probability that the actual samples (**observations**) are less or equal than the predicted values (**red line**) is 70%

Quantile Regression: Prediction Intervals

- By combining **2 Quantile Regressors (QR)**, a **Prediction Interval** can be built using the two sets of predictions produced by these two models.

- We want to compute the **0.9** Prediction Interval (**90%**):
  - Fit a 0.05 QR and a 0.95 QR
  - Use the predictions of the 0.05 QR as **l**ower boundary
  - Use the predictions of the 0.95 QR as **u**pper boundary
  - Knowing that:

$$P(\mathbf{l} \leq X \leq \mathbf{u}) = P(X \leq \mathbf{u}) - P(X \leq \mathbf{l}) = 0.95 - 0.05 = \mathbf{0.9}$$

- This means the prediction interval will contain approximately **90%** of samples!

- PI generic formula:

  with α as the miscoverage level

$$\widehat{C}_\alpha(x_t) = \left[\widehat{q}_{\alpha_{lo}}(x_t)\ ,\ \widehat{q}_{1-\alpha_{hi}}(x_t)\right] \begin{cases} \alpha_{lo} + \alpha_{hi} & = \alpha \\ \alpha_{lo}, \alpha_{hi} & \geq 0 \end{cases}$$

How can we evaluate the interval performance?

Prediction Interval (PI) Metrics

Reliability
PI Coverage

- **ACE (Average Coverage Error)**

ACE = PICP - PINC

PICP (Prediction Interval Coverage Probability)

$$\text{PICP} = \frac{1}{N_p}\sum_{i=1}^{N_p}\rho_i \qquad \rho_i = \begin{cases} 0, & if \quad y_i \notin [L_i, U_i] \\ 1, & if \quad y_i \in [L_i, U_i] \end{cases}$$

PINC (Prediction Interval Nominal Coverage)

PINC = 1−α = 90%

Sharpness
PI Size

- **PINAW (Prediction Interval Normalized Average Width)**

$$PINAW = \frac{1}{NR}\sum_{t=1}^{N}(U_t - L_t)$$

where R = ymax – ymin represents the range of the highest and lowest values of the actual yi

- Interval Score $Sc_t^\alpha = \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}} Sc_t^\alpha(\mathbf{x}_i).$

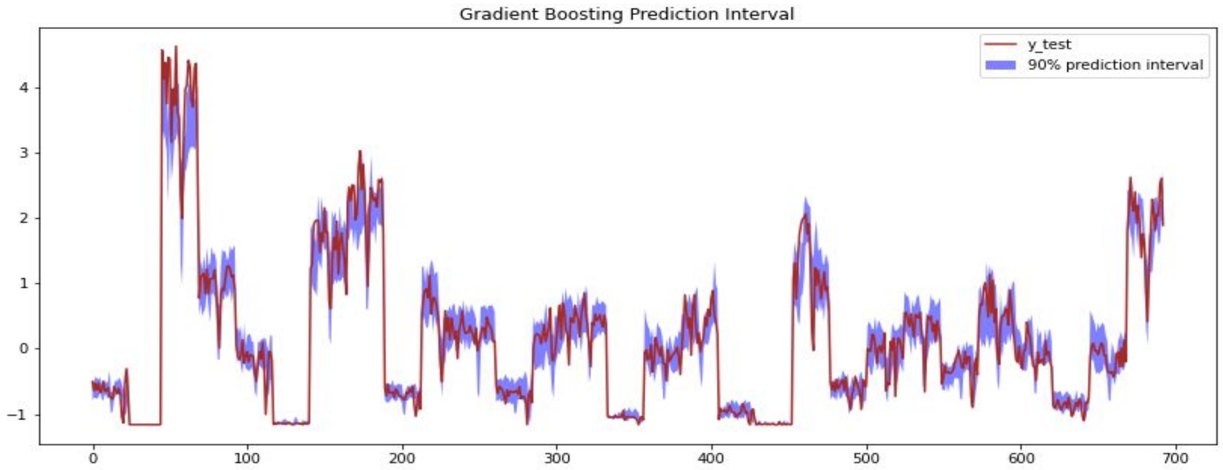
$$Sc_t^\alpha(\mathbf{x}_i) = \begin{cases} -2\alpha\delta_t^\alpha(\mathbf{x}_i) - 4\left[L_t^\alpha(\mathbf{x}_i) - t_i\right], & \text{if } t_i < L_t^\alpha(\mathbf{x}_i) \\ -2\alpha\delta_t^\alpha(\mathbf{x}_i), & \text{if } t_i \in I_t^\alpha(\mathbf{x}_i) \\ -2\alpha\delta_t^\alpha(\mathbf{x}_i) - 4\left[t_i - U_t^\alpha(\mathbf{x}_i)\right], & \text{if } t_i > U_t^\alpha(\mathbf{x}_i). \end{cases}$$

$\delta_t^\alpha(\mathbf{x}_i) = U_t^\alpha(\mathbf{x}_i) - L_t^\alpha(\mathbf{x}_i)$

PINC $100(1-\alpha)\%$

Non-Conformalized PI
Visualization/Metrics

Non-Conformalized Gradient Boosting Quantile Regressor



Target coverage = 90%
Reliability metric:
ACE = 14.1%

Sharpness metric:
PINAW = 0.538

Need of conformalization

Introduction to Conformal Prediction

- Uncertainty quantification methods are able to build Prediction Intervals (PI)
  - but performances are not optimal!

- **Need to calibrate the intervals and guarantee the uncertainty quantification**

- Solution: Conformal Prediction (CP) [Vovk et al, 2005] is a set of:
  - Distribution-free
  - **Model-agnostic**

methods that quantify uncertainty by constructing Prediction Intervals (PI) whose **probability coverage** is **backed by theoretical guarantees**:

With α as the miscoverage level: $$\mathbb{P}\big(Y_{n+1} \in C(X_{n+1})\big) \geq 1 - \alpha$$

Conformal Prediction (CP)
Mean-based vs Quantile-based

CP can be used as layer to:

Compute a PI for a point-based regressor

Mean-based regressor

$$\widehat{C}_\alpha(X) = \left[\, \widehat{f}(X) - \delta^f_\alpha \,,\; \widehat{f}(X) + \delta^f_\alpha \right], \quad \delta^f_\alpha \geq 0$$

Error margin

Adjust the PI of an interval-based regressor

Quantile-based regressor

$$\widehat{C}_\alpha(X) = \left[\widehat{q}_{\alpha_{lo}}(x) - \delta^q_\alpha,\; \widehat{q}_{1-\alpha_{hi}}(x) + \delta^q_\alpha\right], \quad \delta^q_\alpha \in \mathbb{R}$$

Error margin

- Error margin is derived from the (1−α)-th empirical quantile of specifically designed regression residuals, known as **nonconformity scores**
- They are proper to each CP algorithm
- They are computed by evaluating the predictor **on held-out calibration data**

Conformal Prediction (CP)

Data Scheme

Split

Partition
train data into fit
and calibration
[Lei et al, 2018]

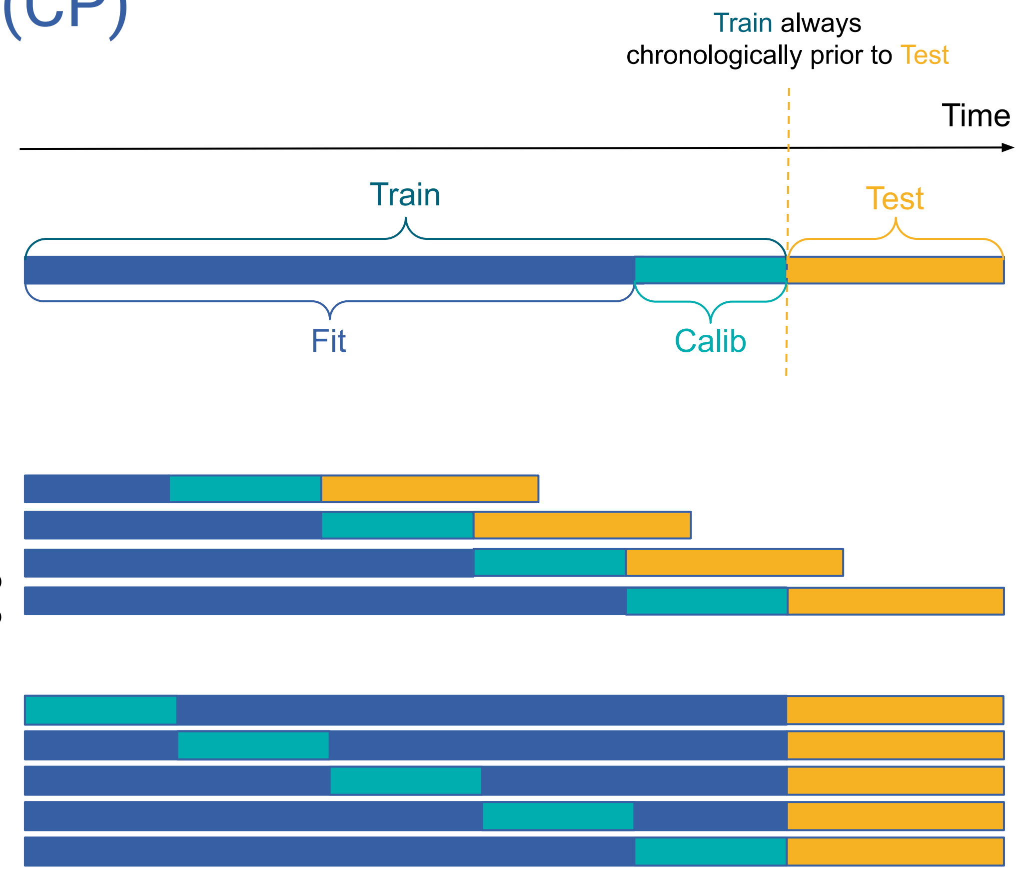
Simplest and
cheapest way

K-fold

Partition data into K folds
and fit K models
[Barber et al, 2021]

Could enhance
calibration
generalization
(when temporal gap
Fit/Calib is similar to
Calib/Test)

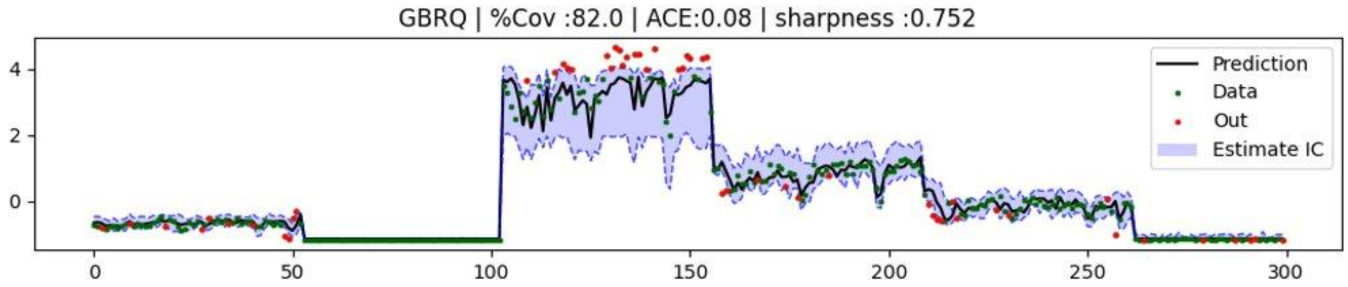
Fit + Calibration
on whole data



Conformalized PI

Metrics/Visualization

Conformalized Gradient Boosting Quantile Regressor



Target coverage = 90%

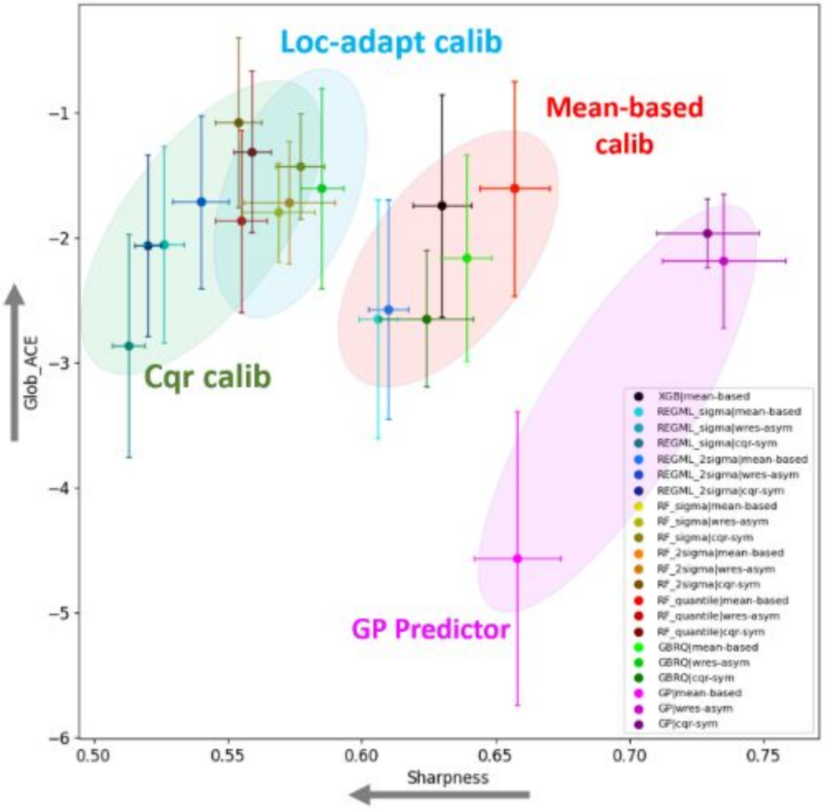
Reliability metric:

ACE = 8%

Sharpness metric:

PINAW = 0.752

Current benchmark results: UQ + CP



Thank you for your attention!