Leveraging conformal prediction

for calibrated probabilistic time series forecasts to accelerate the renewable energy transition

Inge van den Ende
Data Scientist

www.dexterenergy.ai

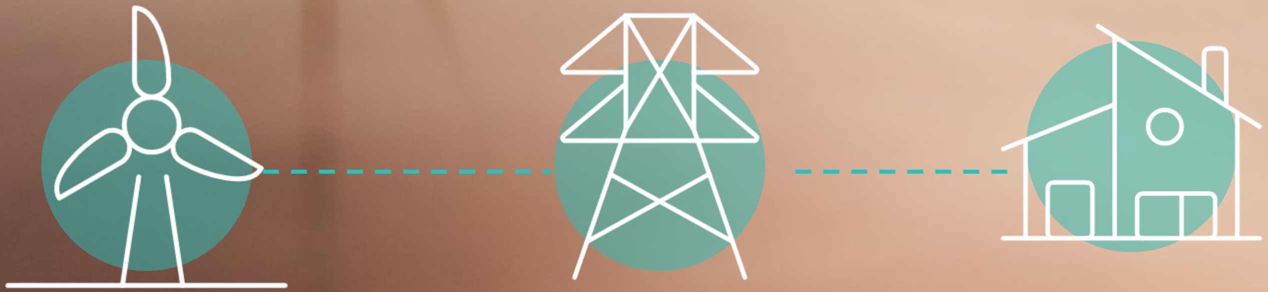
pyladies amsterdam

25 January 2024

DEXTER

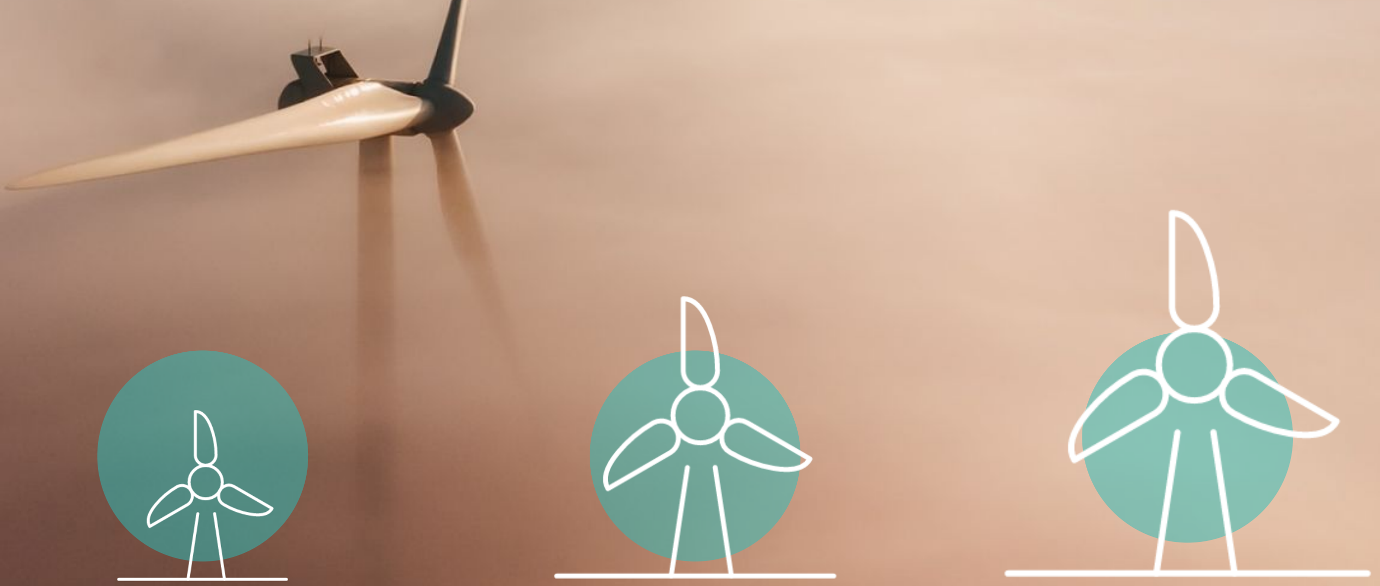
Photo by Nicholas Doherty on Unsplash

A balancing act on the energy grid: Supply needs to equal demand at any moment



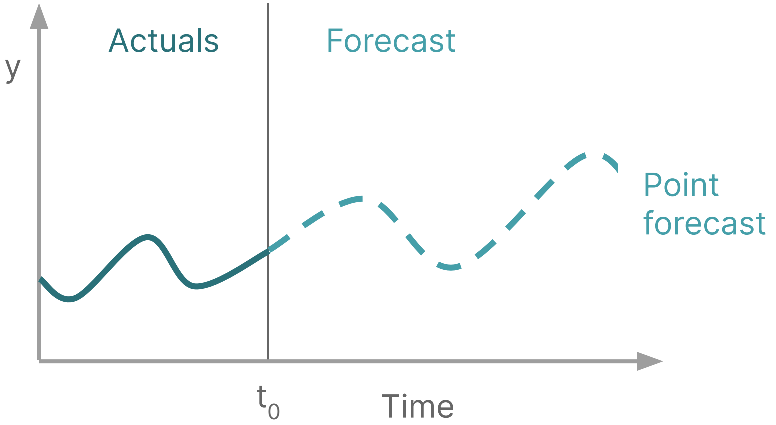
$supply_t$ = $demand_t$

Uncertainty in energy generation forecast increases

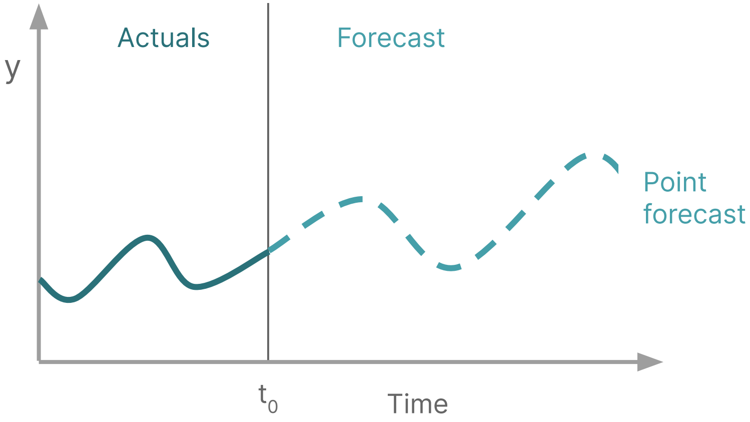


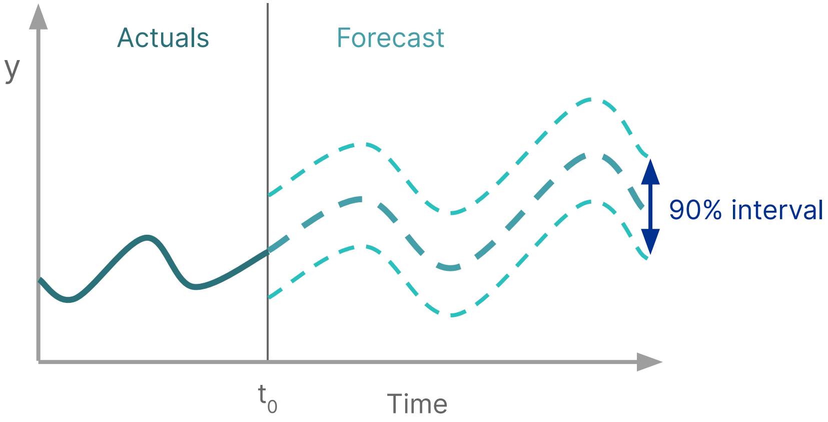
Forecasting the uncertainty explicitly enables decision making

Point forecasts don't give any information about this uncertainty

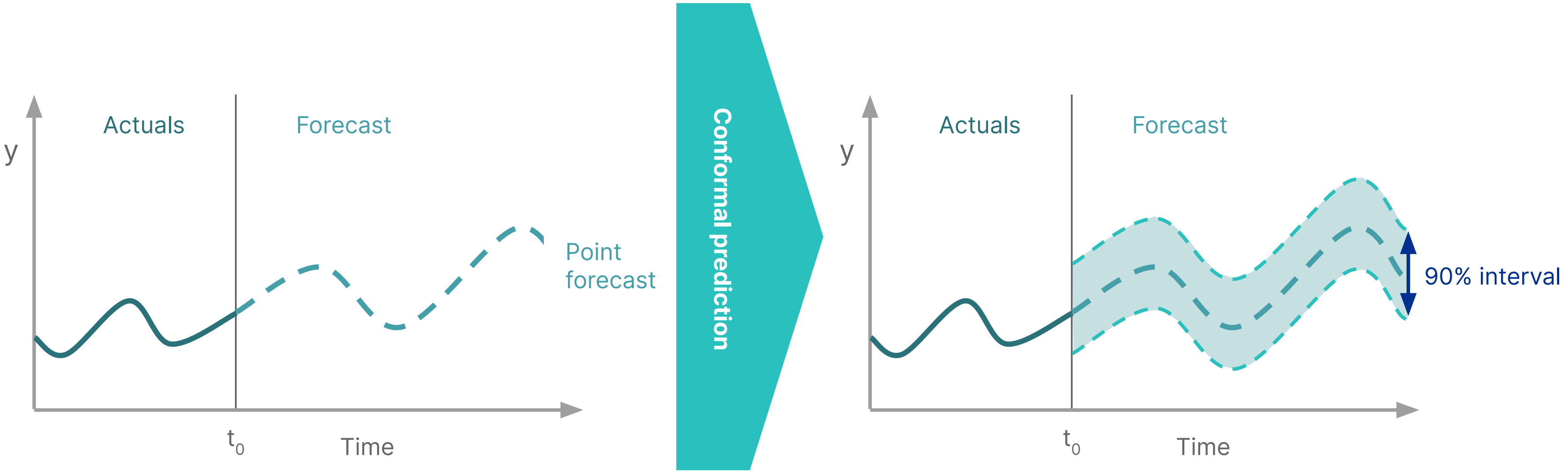


A prediction interval gives us more information about the uncertainty



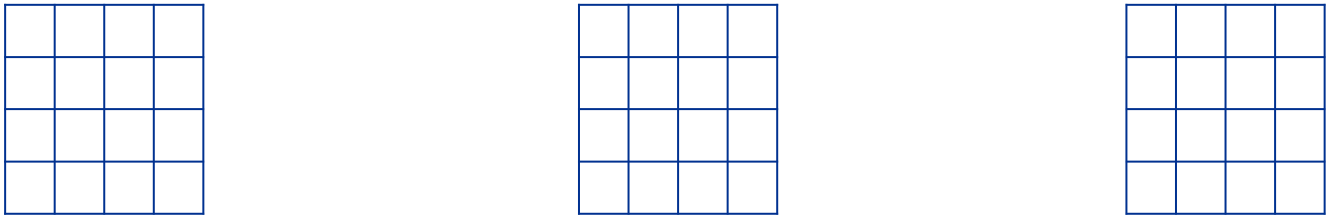


Conformal prediction can create a prediction interval for any point forecast



A calibration set is hold out from the train set

Train set | Calibration set | Test set



Three steps to forecast with prediction interval

Training	Calibration	Prediction

Train the point forecast model on the train set as usual

Training	Calibration	Prediction

1.

Train point forecast model
on the training set

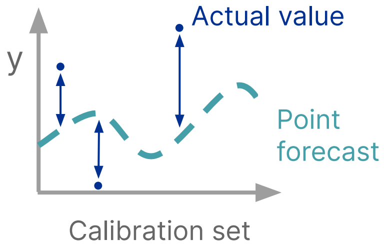
$$\hat{y} = f(X)$$

The calibration set is used to compute the prediction interval

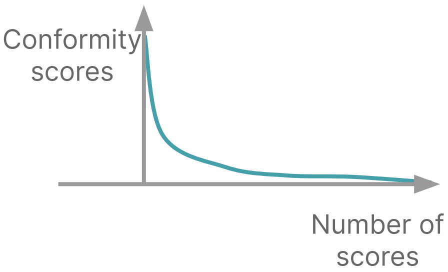
Training | Calibration | Prediction

2.

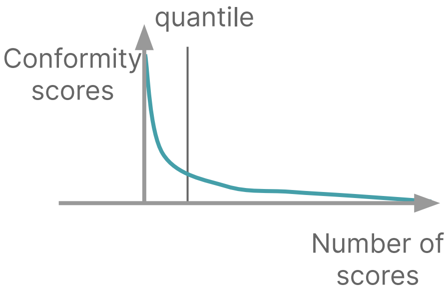
Predict on calibration set and get conformity scores



Create a sorted list with absolute conformity scores



Select the quantile of the prediction interval

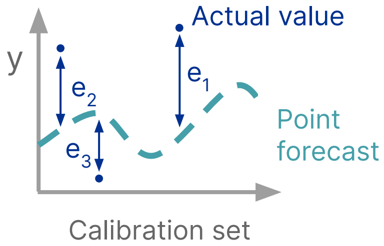


The calibration set is used to compute the prediction interval

Training	Calibration	Prediction

2.

Predict on calibration set and get conformity scores



Create a sorted list with absolute conformity scores

$$\begin{bmatrix} |e|_1 \\ |e|_2 \\ \dots \\ |e|_{N-1} \\ |e|_N \end{bmatrix}$$

Select the quantile of the prediction interval

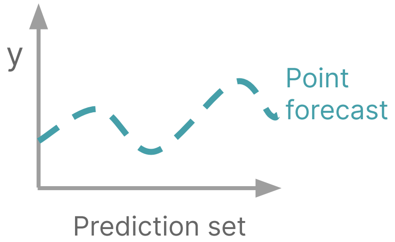
$$\begin{bmatrix} |e|_1 \\ \boxed{|e|_2} \\ \dots \\ |e|_{N-1} \\ |e|_N \end{bmatrix} \quad \text{index} = ((1 - \alpha) * n) - 1$$

Add the prediction interval to every prediction

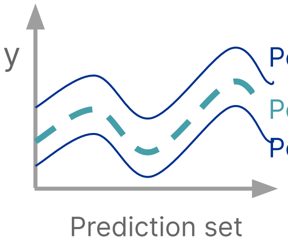
Training | Calibration | Prediction

3.

Predict with point forecast model



Add prediction interval based on conformity scores



Point forecast + selected conformity score $= y + |e|_2$

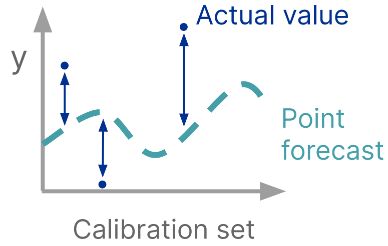
Point forecast $= y$

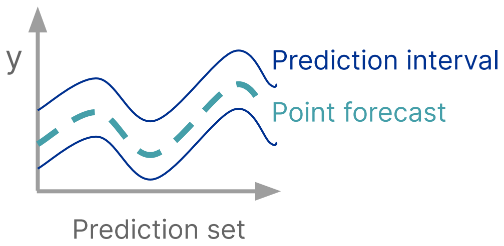
Point forecast - selected conformity score $= y - |e|_2$

Residuals of a calibration set determine the prediction interval

Training	Calibration	Prediction
1.	**2.**	**3.**
Train point forecast model	Predict on calibration set and get conformity scores	Predict with point forecast model & add prediction interval based on conformity scores

$$\hat{y} = f(X)$$





Python packages for conformal prediction

MAPIE:

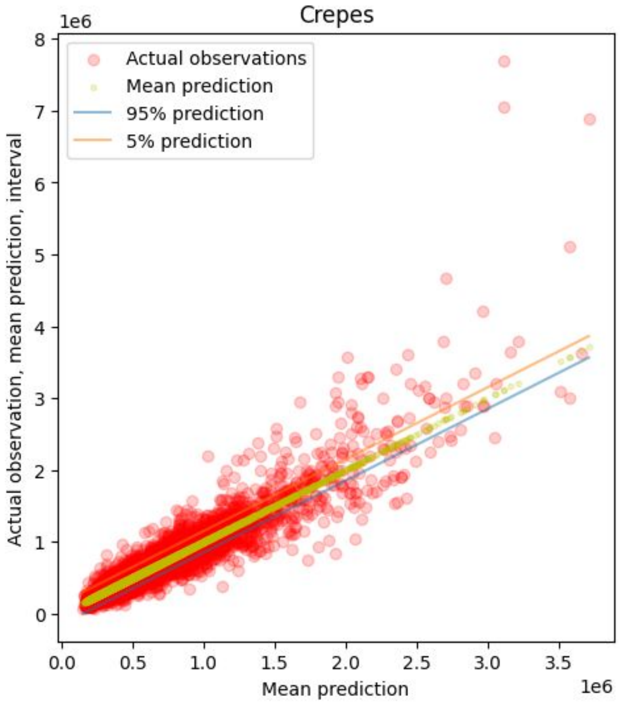
Model Agnostic Prediction Interval Estimator

Crepes:

Conformal classifiers, regressors and predictive systems

Forecasting with prediction interval with the crepes package

- crepes_model = WrapRegressor(baseline_model)
- crepes_model.fit(X_prop_train, y_prop_train)
- crepes_model.calibrate(X_cal, y_cal)
- crepes_point_prediction = crepes_model.predict(X_test)
- crepes_prediction_cp = crepes_model.predict_int(X_test, confidence=0.90)



This simple method has great advantages, but also some disadvantages

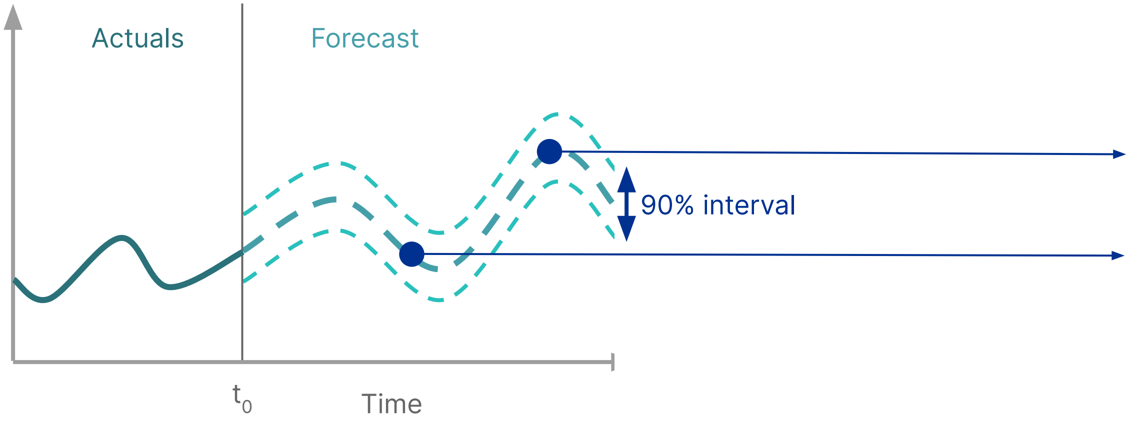
Advantages

- ✓ Model agnostic: Any model can be used
- ✓ Statistical guarantee: valid coverage
- ✓ No distribution assumption needed

Disadvantages

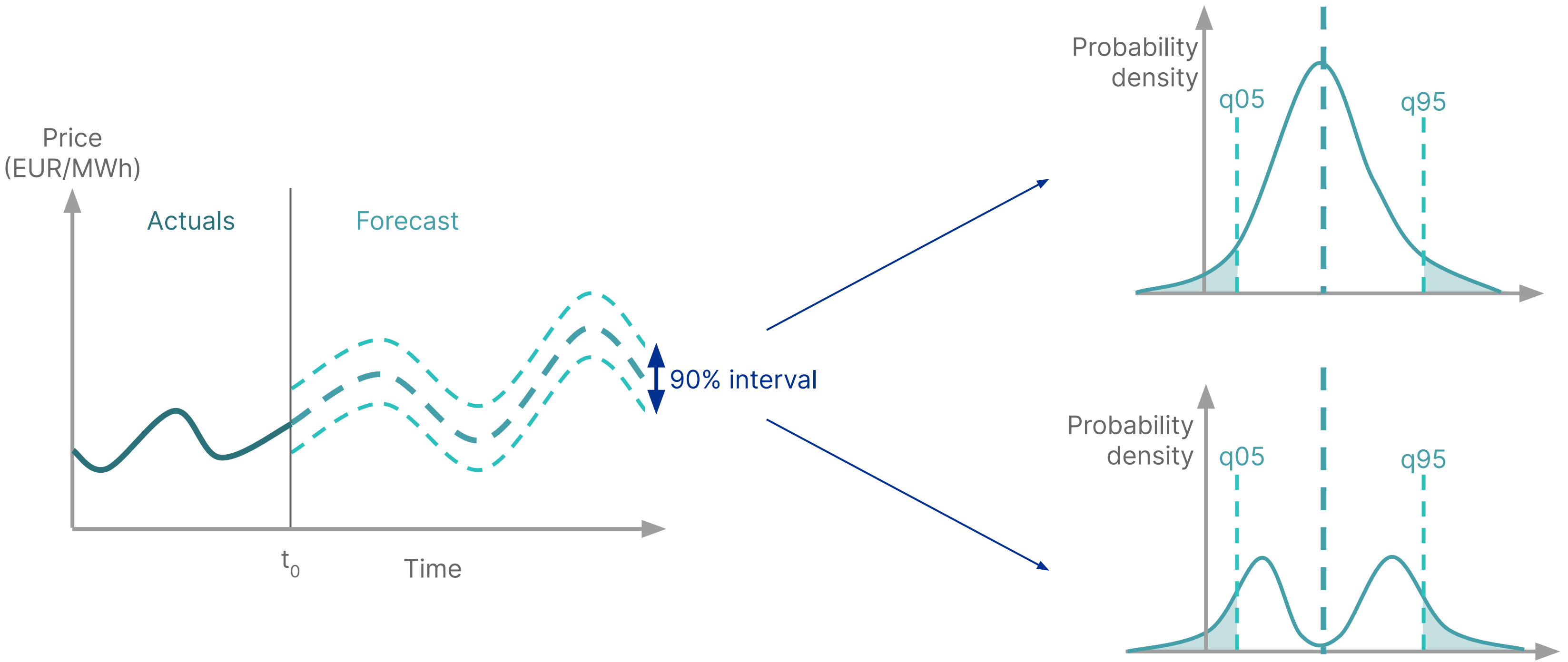
- ✗ Constant over the prediction set
- ✗ A single prediction interval provides less information then a distribution

The prediction interval is constant over the prediction set



Do we expect the same uncertainty for these points?

A prediction interval provides less information then a probabilistic distribution



This simple method has great advantages, but also some disadvantages

Advantages

- √ Model agnostic: Any model can be used
- √ Statistical guarantee: valid coverage
- √ No distribution assumption needed

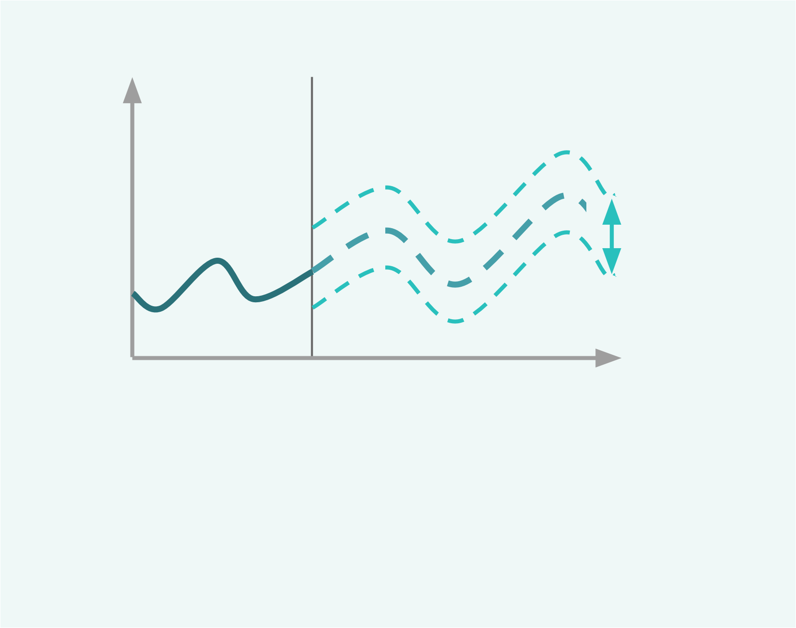
Disadvantages

- X Constant over the prediction set
- X A single prediction interval provides less information then a distribution

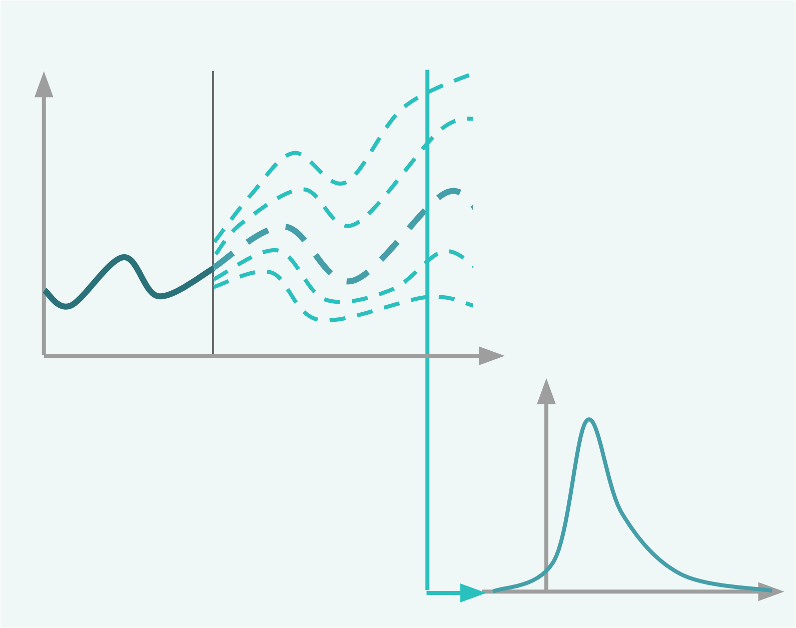
Solution will be given in next slides

Calibrating a probabilistic forecast creates a well-calibrated full distribution that is specific over samples

Part 1: Create prediction interval



Part 2: Calibrate probabilistic forecast



We can use the same three steps to calibrate a probabilistic forecast

Training	Calibration	Prediction

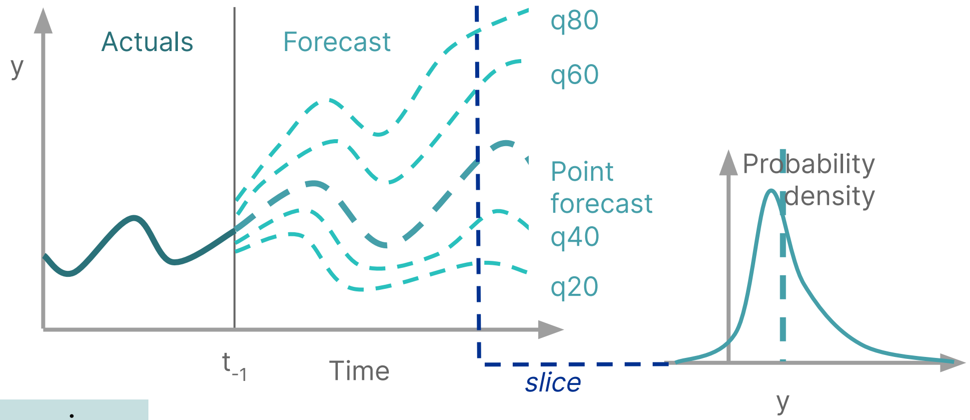
Train a probabilistic forecast model on the train set

Training	Calibration	Prediction

1.

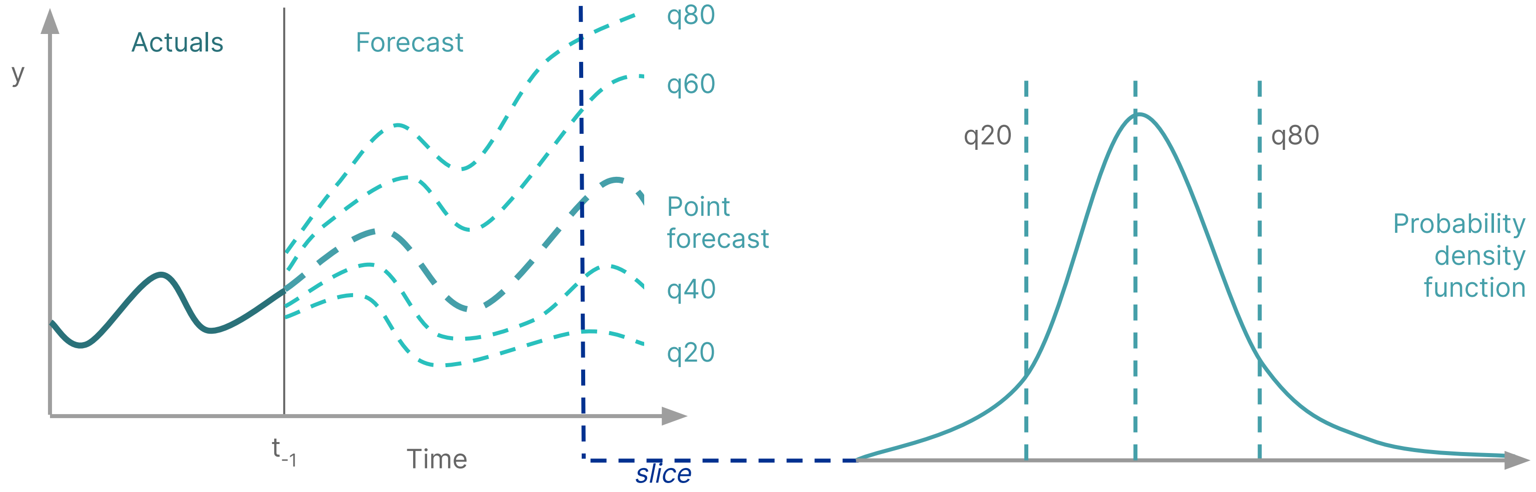
Train a probabilistic forecast model on the training set

$$[\hat{y}_{q01} \ldots \hat{y}_{q01}] = f(X)$$

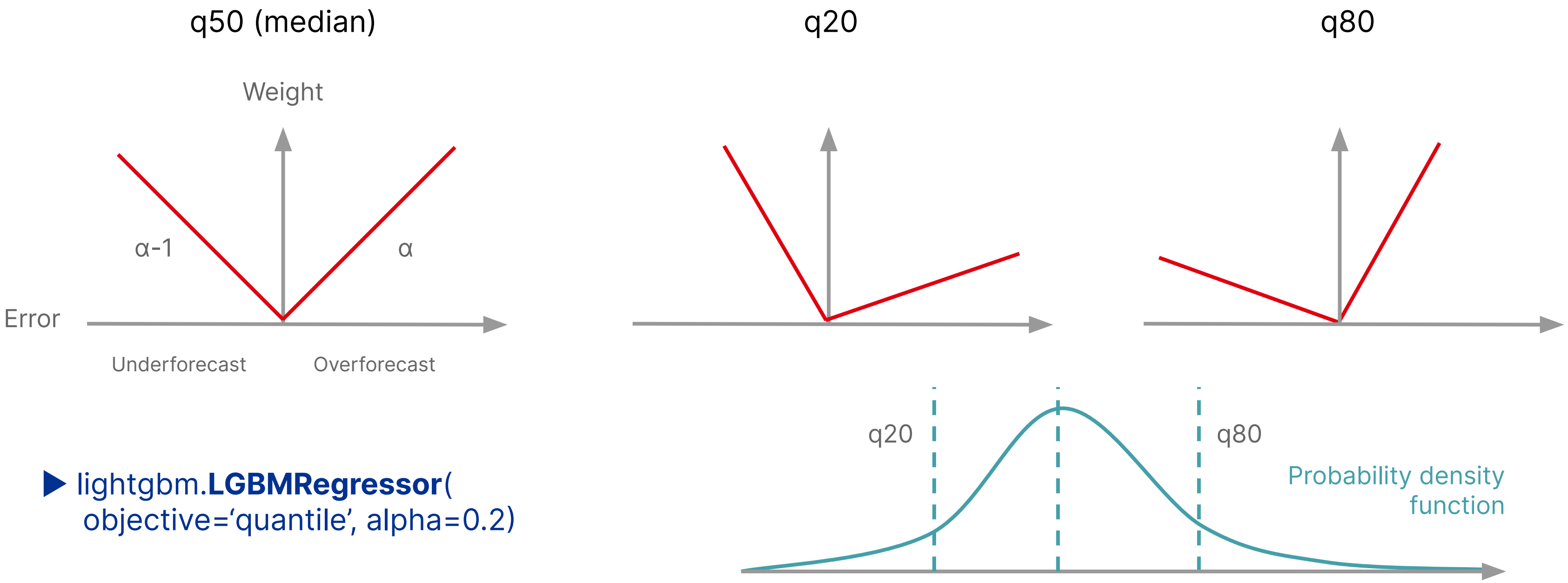


For example: conformalized quantile regression

Quantile regression: fit a model per quantile that you predict



Quantile regression: asymmetrically weight errors during model training



- lightgbm.**LGBMRegressor**(objective='quantile', alpha=0.2)

Why do we need conformalized quantile regression?

Quantile regression

- ✗ Asymptotically consistent
- ✓ Takes into account local variability of the input space

Conformal prediction

- ✓ Statistical guarantee of valid coverage
- ✗ Basic application does not adapt to input space

Conformalized quantile regression

- ✓ Statistical guarantee of valid coverage
- ✓ Takes into account local variability of the input space

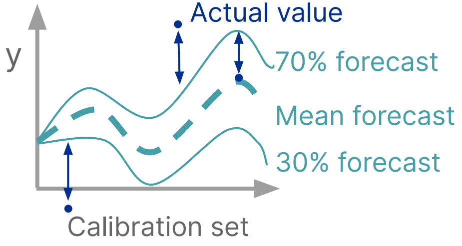
The calibration set is used to compute a correction factor

Training | Calibration | Prediction

Per interval of the probabilistic prediction

2.

Predict on calibration set and get conformity scores



Create a sorted list with the conformity scores = residuals

$$\begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_{N-1} \\ e_N \end{bmatrix}$$

Select the quantile of the prediction interval

$$\begin{bmatrix} e_1 \\ \boxed{e_2} \\ \dots \\ e_{N-1} \\ e_N \end{bmatrix}$$

index = $((1 - \alpha) * n) - 1$

Correction can be positive or negative:
Positive = wider distribution
Negative = more narrow

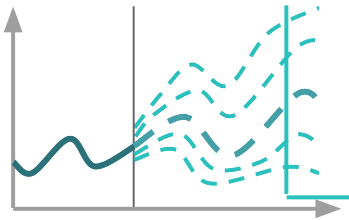
Calibrate every prediction interval

Training | Calibration | Prediction

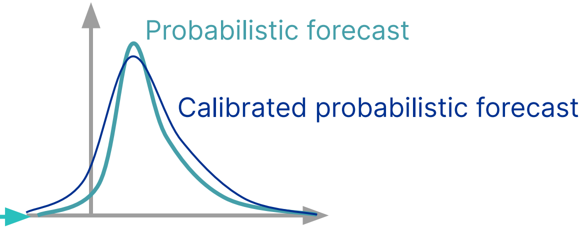
Per interval of the probabilistic prediction

3.

Predict with probabilistic forecast model



Calibrate distribution based on conformity scores

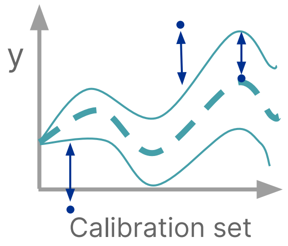


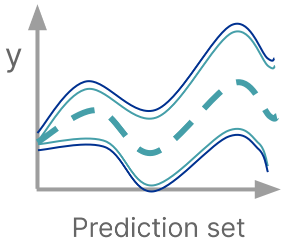
DEXTER

Residuals of a calibration set are used to calibrate the forecasted distribution

Training	Calibration	Prediction
1.	**2.**	**3.**
Train probabilistic forecast model	Predict on calibration set and get conformity scores for every quantile	Predict on test set and calibrate that distribution

$$[\hat{y}_{q01} \ldots \hat{y}_{q01}] = f(X)$$





A remaining disadvantage: exchangeability

Advantages

√ Model agnostic: Any model can be used

√ Statistical guarantee: valid coverage

√ No distribution assumption needed

Advantages

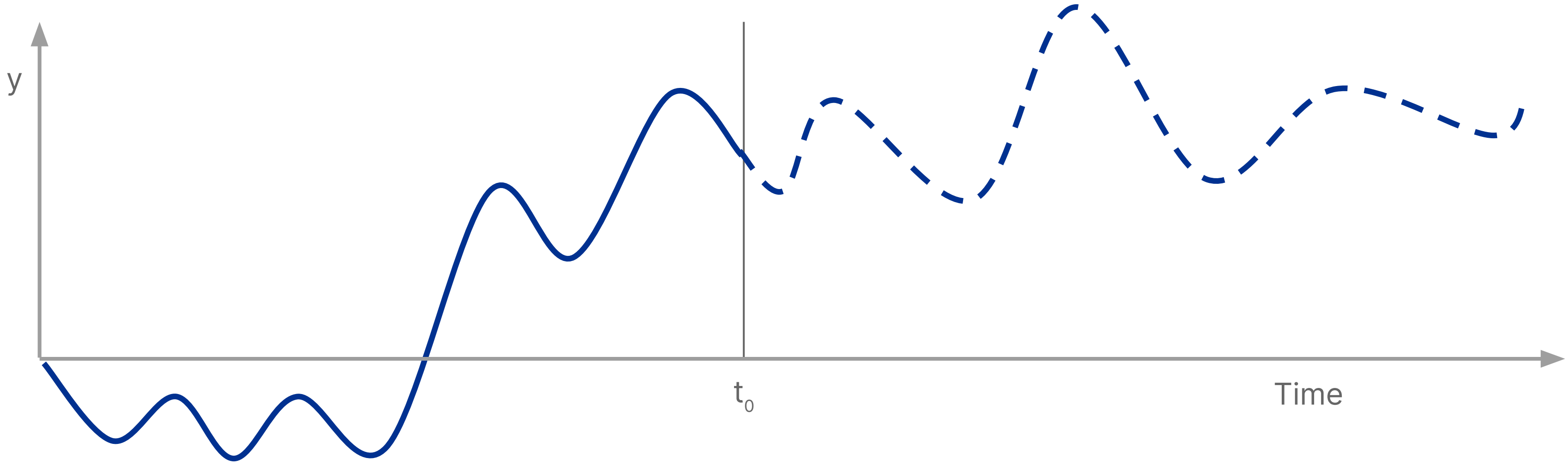
√ Varies over the prediction set

√ A distribution provides more information then a single prediction interval

Disadvantage

Assumption: exchangeability

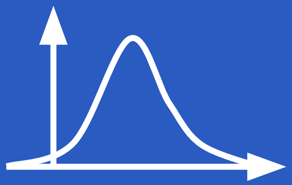
Exchangeability does not always hold



▶ mapie.time_series_regression.MapieTimeSeriesRegressor

Key takeaways about conformal prediction

Simple method with
statistical guarantee

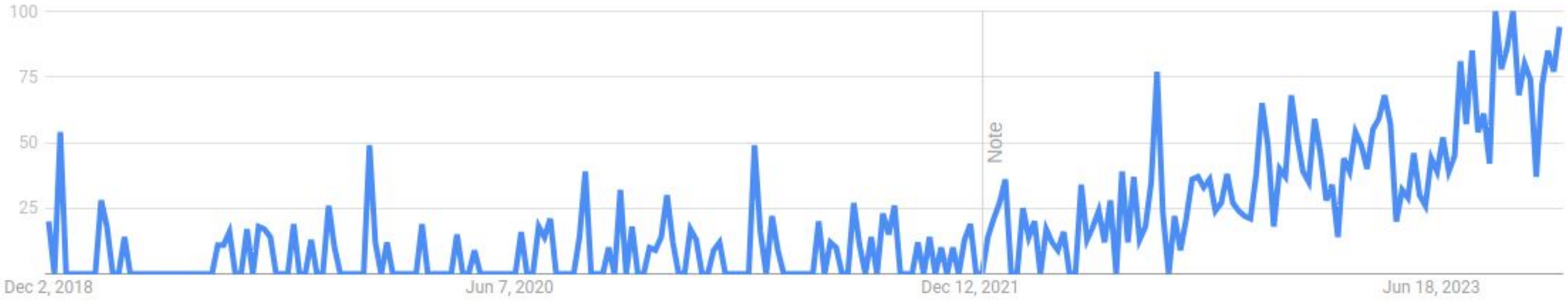


Conditional when calibrating
probabilistic forecast

Helps to accelerate the
renewable energy transition

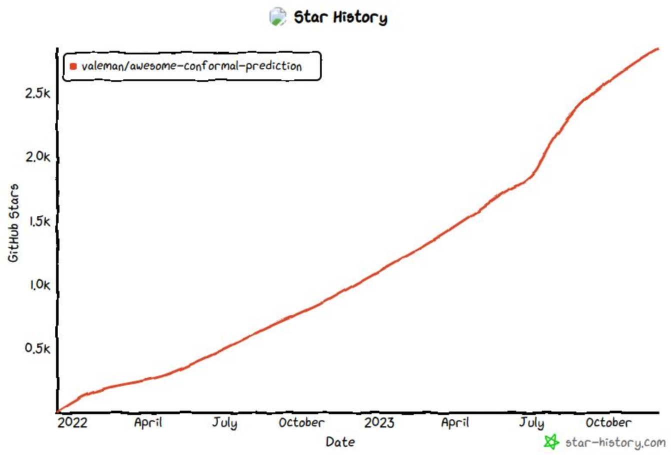
At the start of 2022 the interest in conformal prediction started to rise

Google trend worldwide show increase from start of 2022



On the awesome-conformal-prediction github you can find more information

Started in 2022



QR code to awesome-conformal-prediction github

Thank you

DEXTER

www.dexterenergy.ai

Photo by Nicholas Doherty on Unsplash