International Institute of Forecasters

Forecasting for Social Good · Since 2020 ·

# *Probabilistic Forecating in Python*

Presented by

**Sankalp Gilda, Ph.D.**

**Aug 2024**

# Content



August 2024

Forecasting for Social Good
Since 2020

## Why Time Series?

*"The amount of time series data being stored across organizations has increased 15x over the past 3 years."*

--Timescale, "The State of Time Series Data 2023" report

*"By 2027, 28.8% of the Global DataSphere will be real-time data."*

-- IDC, "Global DataSphere Forecast, 2023–2027"
(Published in May 2023)

*"78% of respondents report that the amount of time series data they are managing has increased over the past 12 months."*

-- InfluxData's "State of Time Series Data Management Report" for 2022

## What is Forecasting?

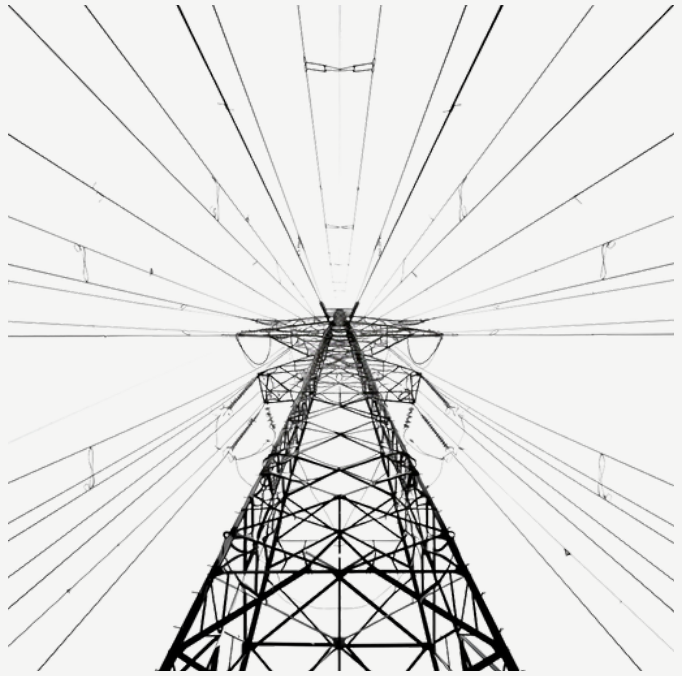
Forecasting is the process of making predictions about future events based on past and present time series data.

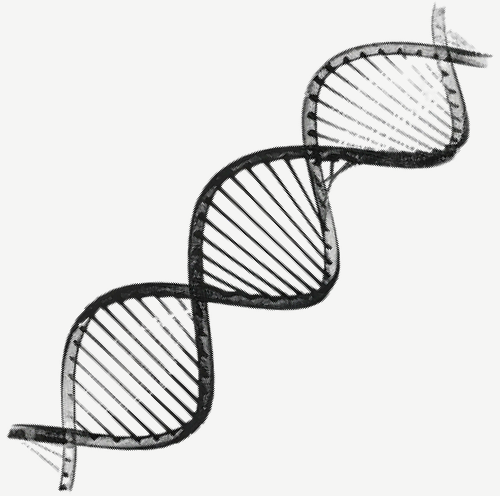
- Uses historical data and current trends
- Applies to various fields (e.g., weather, economics, business)
- Combines science, data analysis, and intuition
- Essential for planning and decision-making

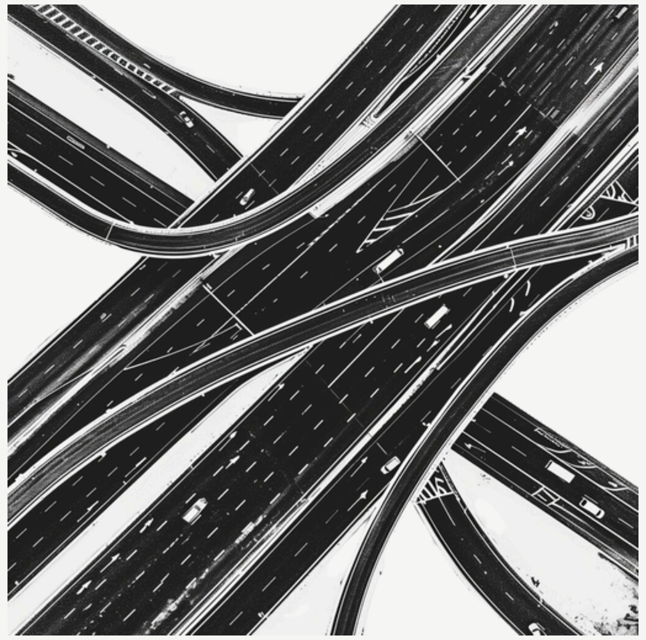

*Remember: All forecasts come with inherent uncertainty*

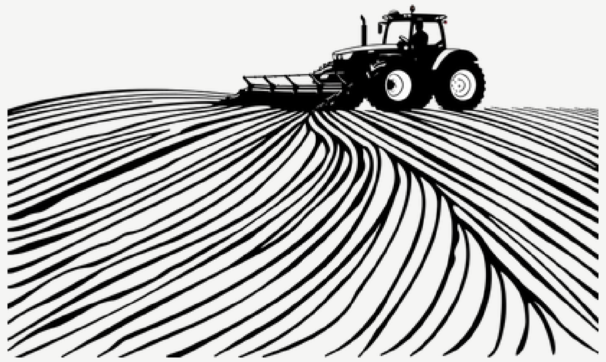
## Why Forecasting?

Forecasting for Social Good
Since 2020









# Why Forecasting?

Forecasting for Social Good
Since 2020

**Weather and Climate**

Forecasting for Social Good
Since 2020

# Why Forecasting?

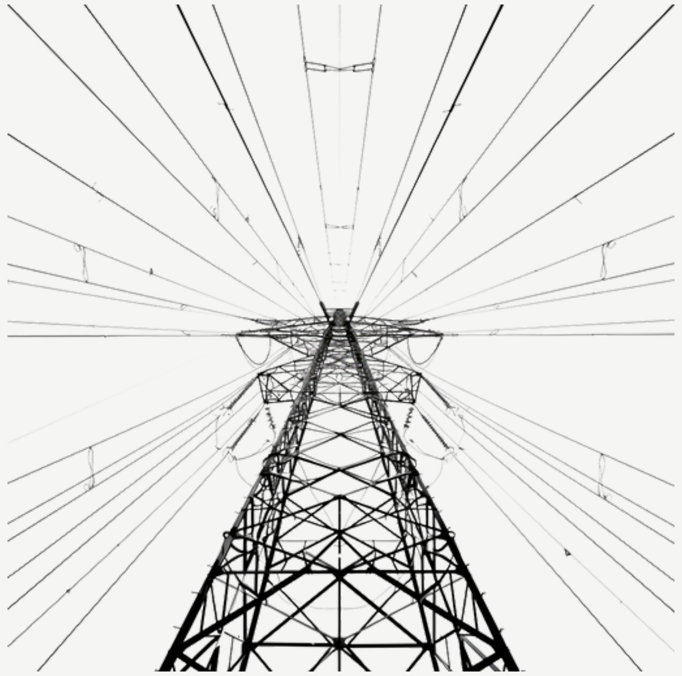
Weather and Climate

Finance and Economics

# Why Forecasting?

Weather and Climate

Finance and Economics



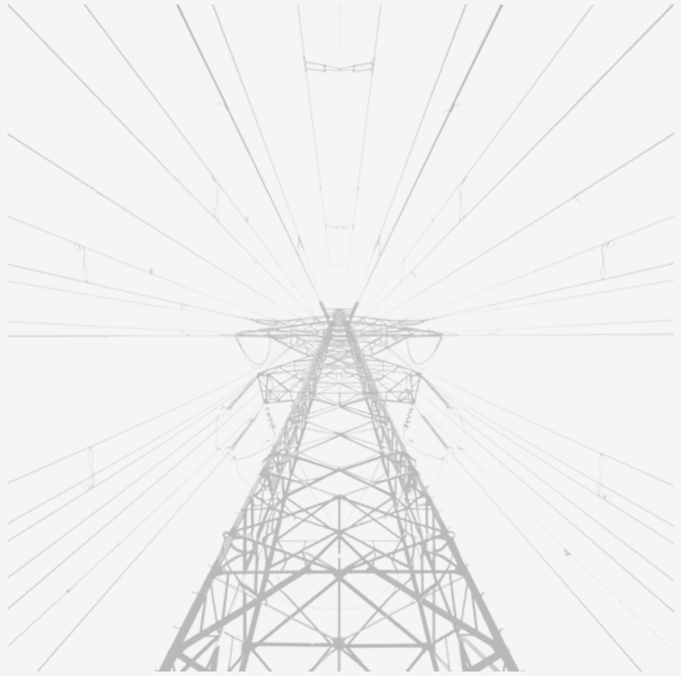
Energy

Forecasting for Social Good
Since 2020

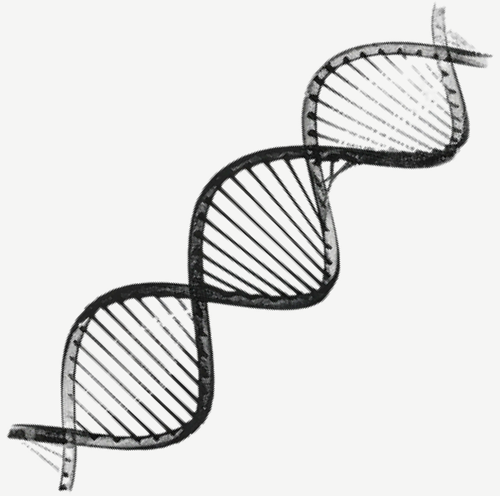
# Why Forecasting?

Weather and Climate

Finance and Economics



Energy



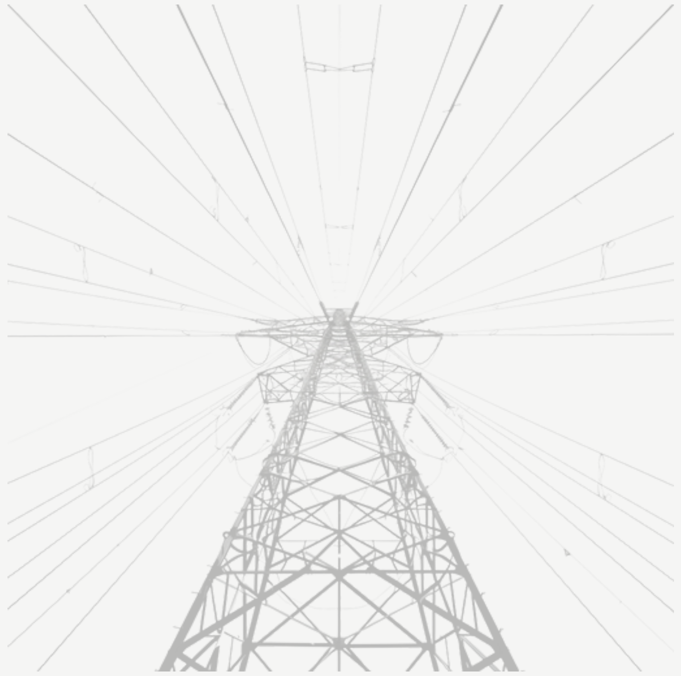
**Healthcare**

Forecasting for Social Good
Since 2020

## Why Forecasting?

Weather and Climate

Finance and Economics



Energy

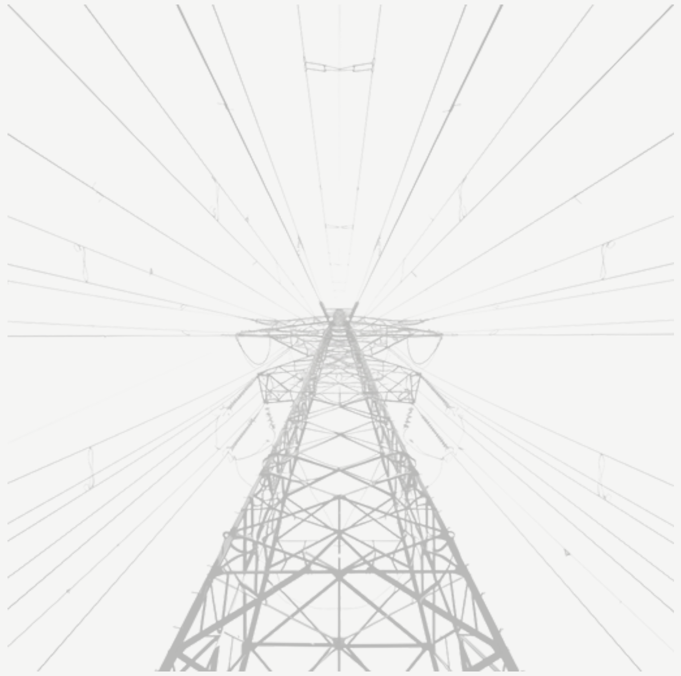
Healthcare

**Retail and E-Commerce**

## Why Forecasting?

Weather and Climate

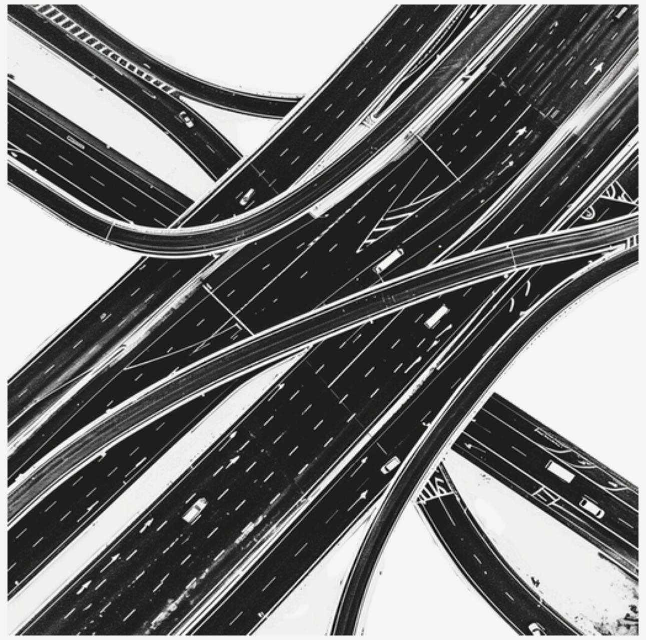
Finance and Economics



Energy

Healthcare

Retail and E-Commerce



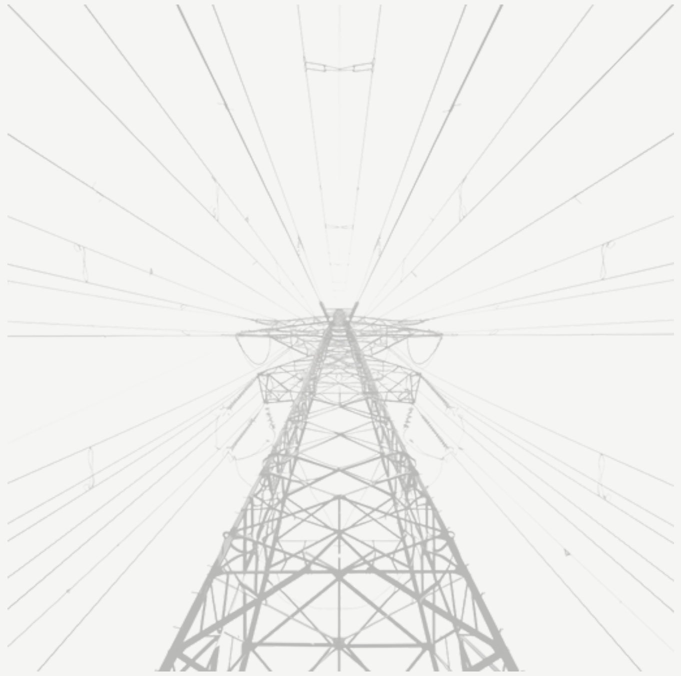
**Transportation and Logistics**

Forecasting for Social Good
Since 2020

# Why Forecasting?


Weather and Climate


Finance and Economics


Energy


Healthcare


Retail and E-Commerce


Transportation and Logistics

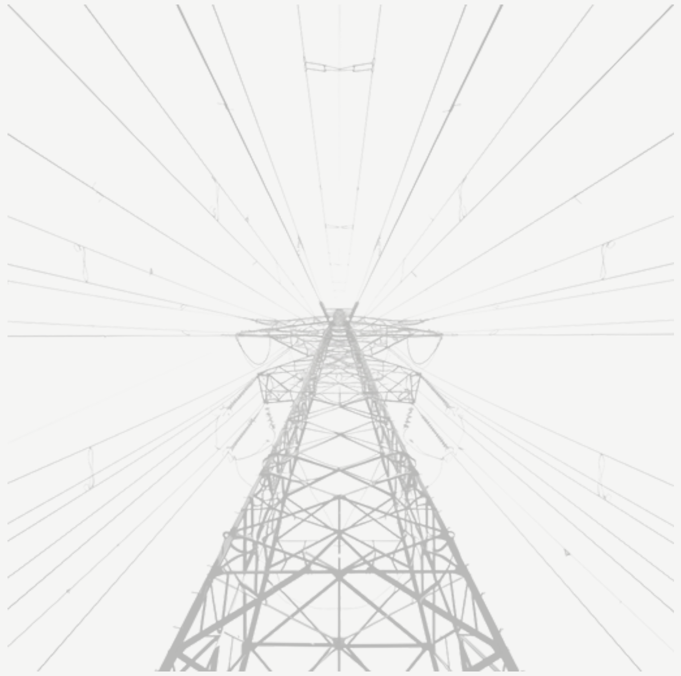

Disaster Management

Forecasting for Social Good
Since 2020

# Why Forecasting?


Weather and Climate


Finance and Economics

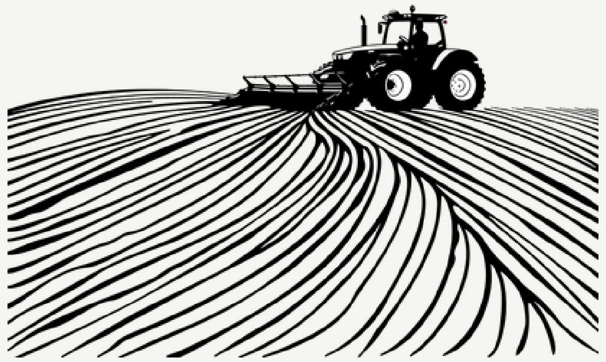

Energy


Healthcare


Retail and E-Commerce

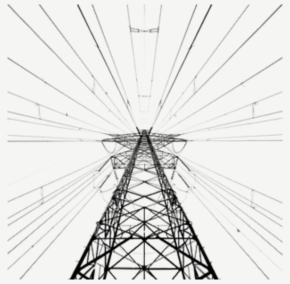

Transportation and Logistics


Disaster Management


**Agriculture**

Forecasting for Social Good
Since 2020

## Why Probabilistic Forecasting?



| | | | |
|---|---|---|---|
| **Scenario A** | Sunny weekend forecast | GDP growth projection | Summer peak demand |
| **What ifs A** | 1. 30% chance of rain?<br>2. Hourly precipitation probability? | 1. Range of GDP growth scenarios?<br>2. Likelihood of recession? | 1. Probability distributions for peak demand?<br>2. Likelihood of extreme heat waves? |
| **Scenario B** | Seasonal winter prediction | Company sales forecast | Wind farm energy production |
| **What ifs B** | 1. Probability distributions for winter severity?<br>2. Likelihood of extreme cold snaps? | 1. Probability distributions for sales outcomes?<br>2. Likelihood of market disruptions? | 1. Probability ranges for wind speeds?<br>2. Likelihood of prolonged low-wind periods? |

## Why Probabilistic Forecasting?

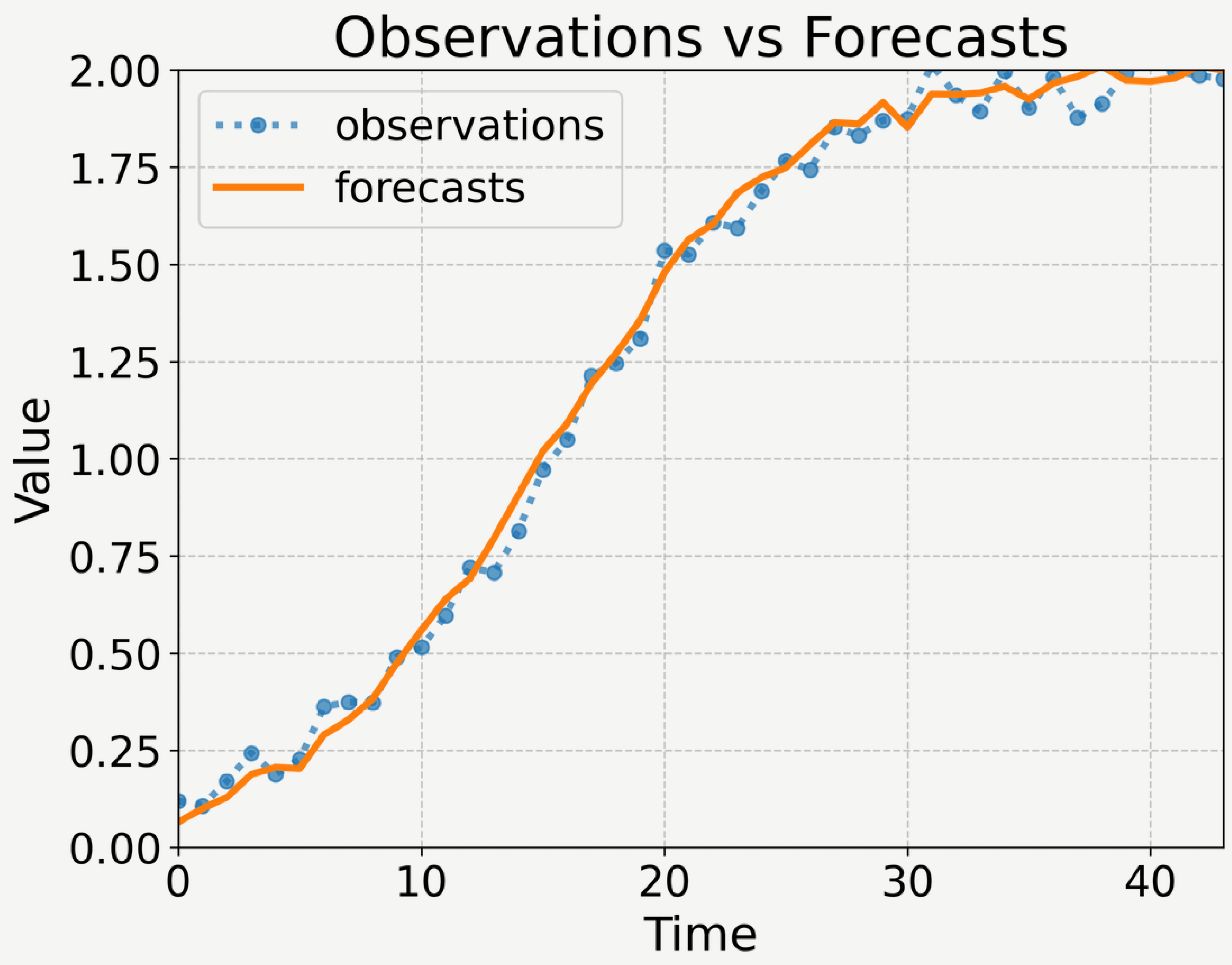
## Point Forecasts

A point forecast is a single-value prediction representing the most likely future outcome, based on current data and models.

$$\hat{y}_{t+k|t} = \mathbb{E}[Y_{t+k}|\mathcal{M}, \Omega_t, \boldsymbol{\theta}_t]$$

where:

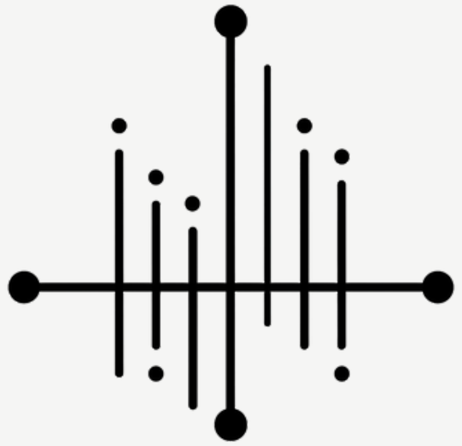
$$\begin{aligned} \hat{y}_{t+k|t} &: \text{forecast for time } t+k \text{ made at time } t \\ \mathbb{E}[\cdot] &: \text{expected value} \\ Y_{t+k} &: \text{actual value at time } t+k \\ \mathcal{M} &: \text{forecasting model} \\ \Omega_t &: \text{information set available at time } t \\ \boldsymbol{\theta}_t &: \text{model parameters estimated at time } t \end{aligned}$$



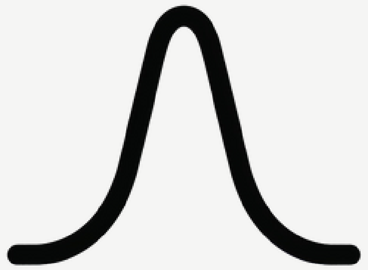
Forecasting for Social Good
Since 2020

# *Types of Probabilistic Forecasts*

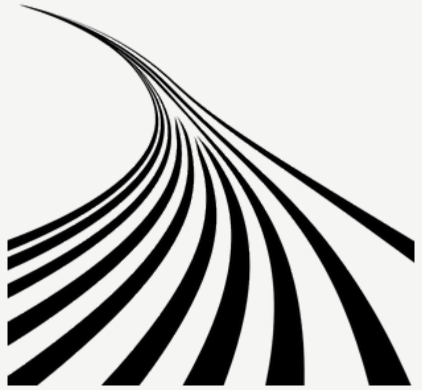
Interval



Quantile



Distribution



Scenario

## Interval Forecasts

A prediction interval is an interval within which power generation may lie, with a certain probability.

$$\hat{I}^{(\alpha)}_{t+k|t} = [\hat{s}^{(q=\alpha/2)}_{t+k|t}, \hat{s}^{(q=1-\alpha/2)}_{t+k|t}]$$

where:

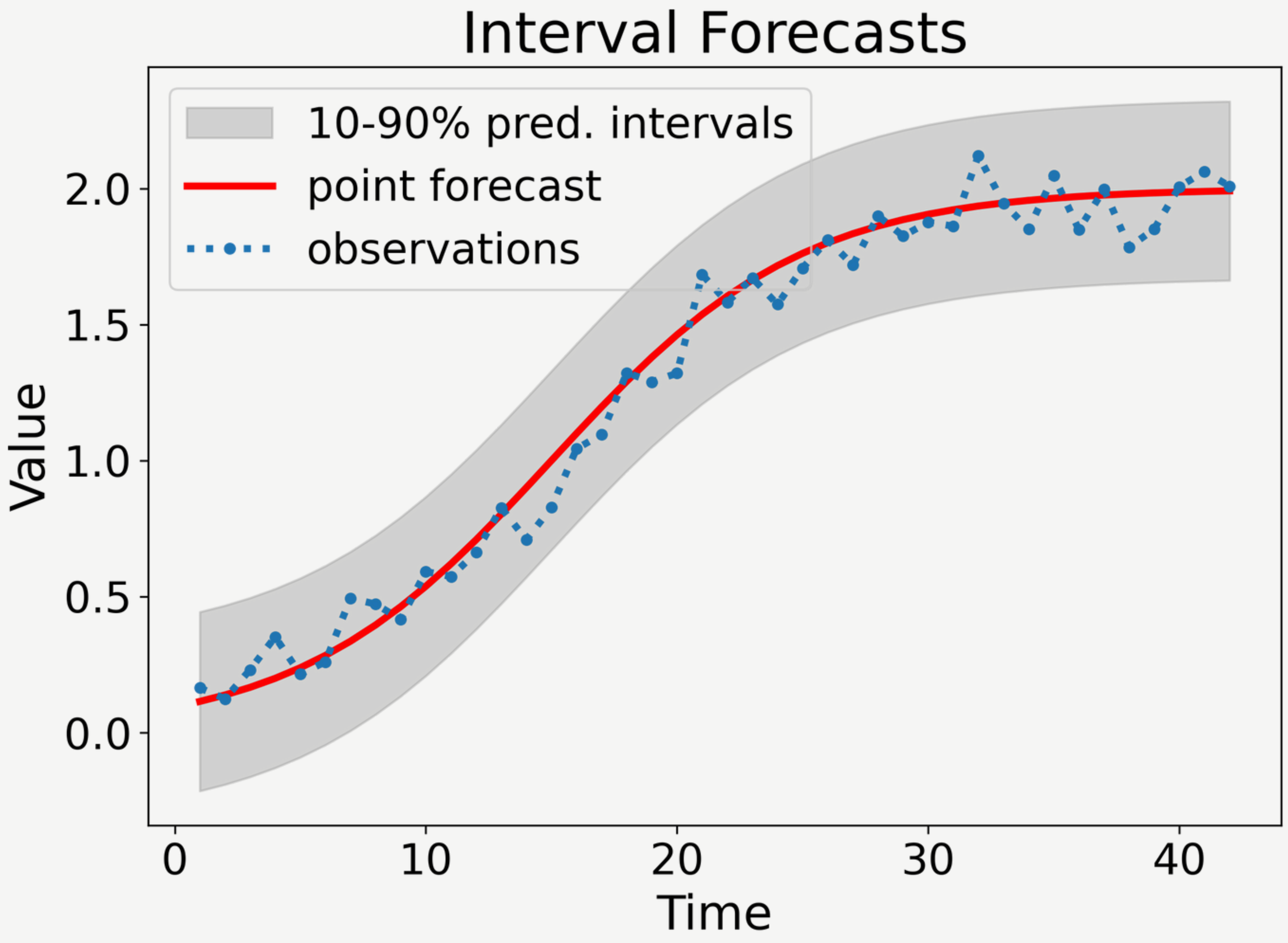
$\hat{I}^{(\alpha)}_{t+k|t}$ : prediction interval

$\hat{s}^{(q)}_{t+k|t}$ : quantiles of the predictive distribution

$\alpha$ : nominal coverage (e.g., 0.9 for 90% interval)

$t+k$ : forecast horizon

$t$ : current time



## Quantile Forecasts



A quantile forecast provides a value that the future observation is expected to be below with a specified probability.

$$P[Y_{t+k|t} \leq \hat{s}^{(q)}_{t+k|t} | g, \Omega_t, \hat{\Theta}_t]$$

$$= q\hat{s}^{(q)}_{t+k|t} = \hat{F}^{-1}_{t+k|t}(q)$$

where:

$Y_{t+k|t}$ : random variable representing the future observation

$\hat{s}^{(q)}_{t+k|t}$ : quantile forecast

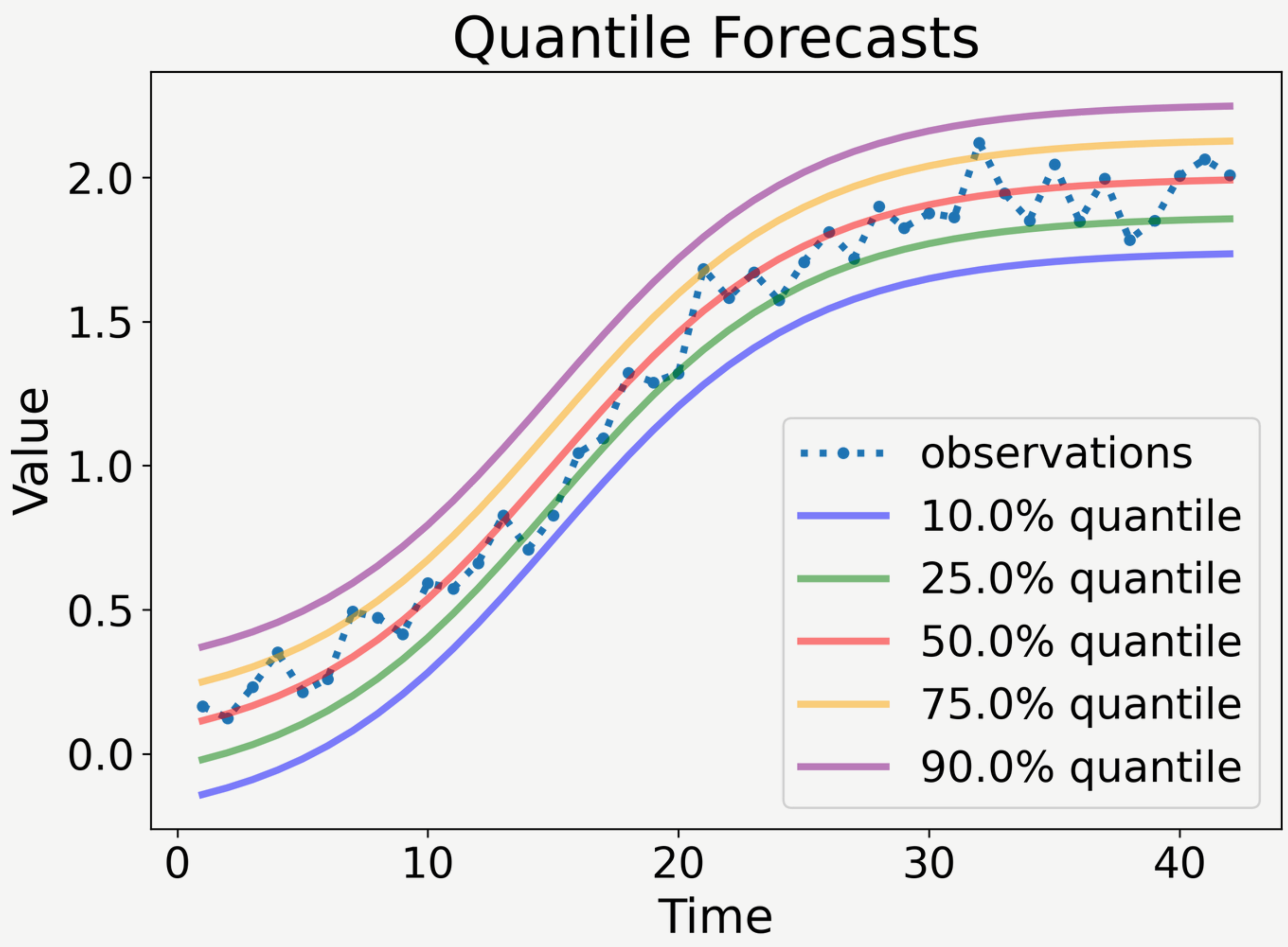
$q$ : nominal level (quantile)

$g$ : forecasting model

$\Omega_t$ : information set available at time $t$

$\hat{\Theta}_t$ : estimated model parameters at time $t$

$\hat{F}_{t+k|t}$ : estimated cumulative distribution function (CDF)



## Distribution Forecasts

A comprehensive probabilistic forecast capturing the full range of potential outcomes across all time horizons.

$$\Omega_{t+k} \sim \hat{\Phi}_{t+k|t}$$

where:

$\Omega_{t+k}$ : Future value at time $t+k$

$\hat{f}_{t+k|t}$ : Estimated predictive density function for time $t+k$ given information up to time $t$

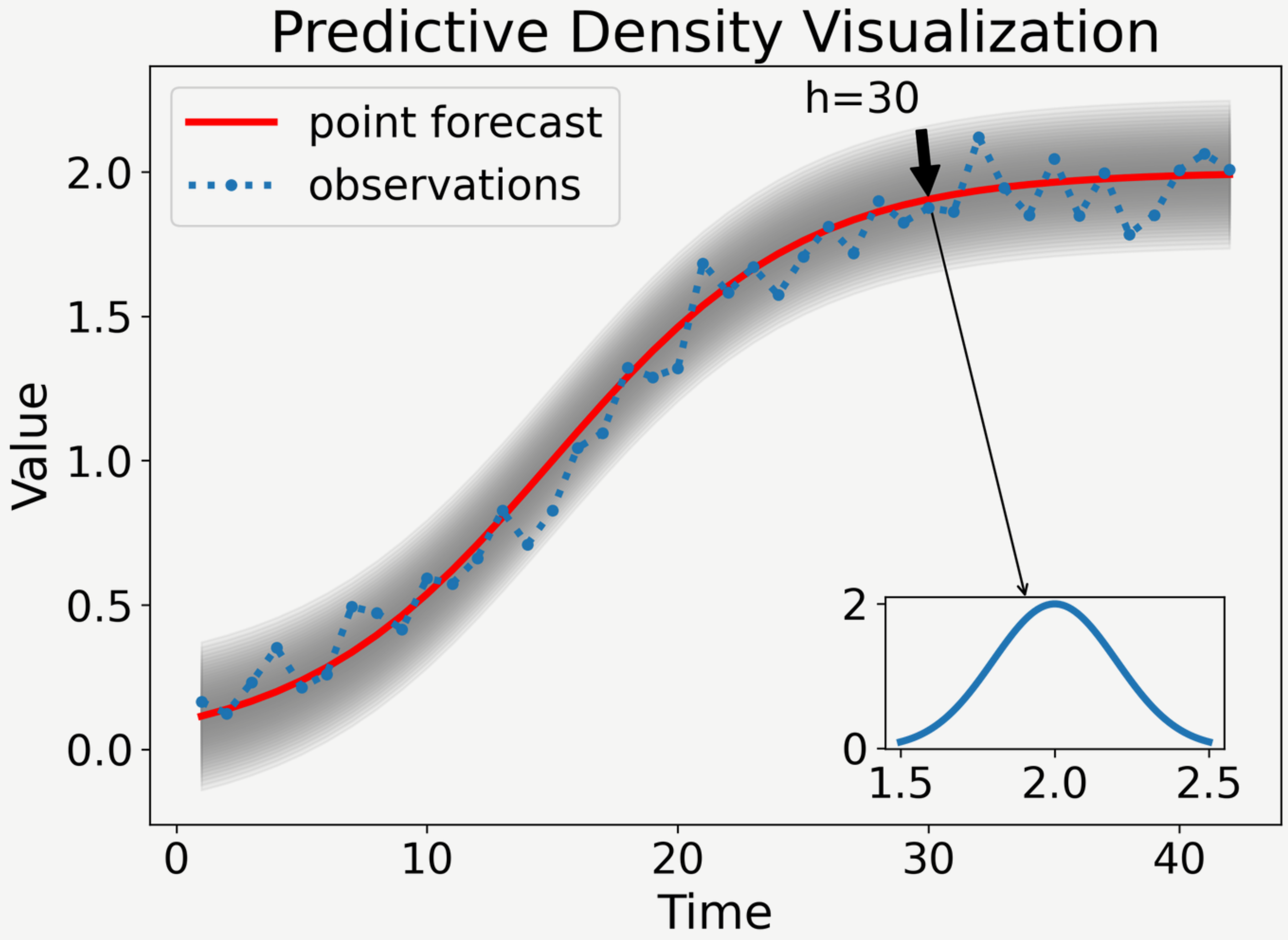
$t$ : Current time

$k$ : Forecast horizon

$\sim$ : denotes "distributed according to"

and:

$$\hat{\Phi}_{t+k|t}(x) = P(\Omega_{t+k} \leq x \mid \text{information up to time } t) = \int_{-\infty}^{x} \hat{f}_{t+k|t}(u)\, du$$



## Scenario Forecasts

A spectrum of potential futures derived from probabilistic modeling to inform decision-making under uncertainty.

$$\boldsymbol{\xi}^{(i)} \sim \hat{\boldsymbol{\Psi}}$$

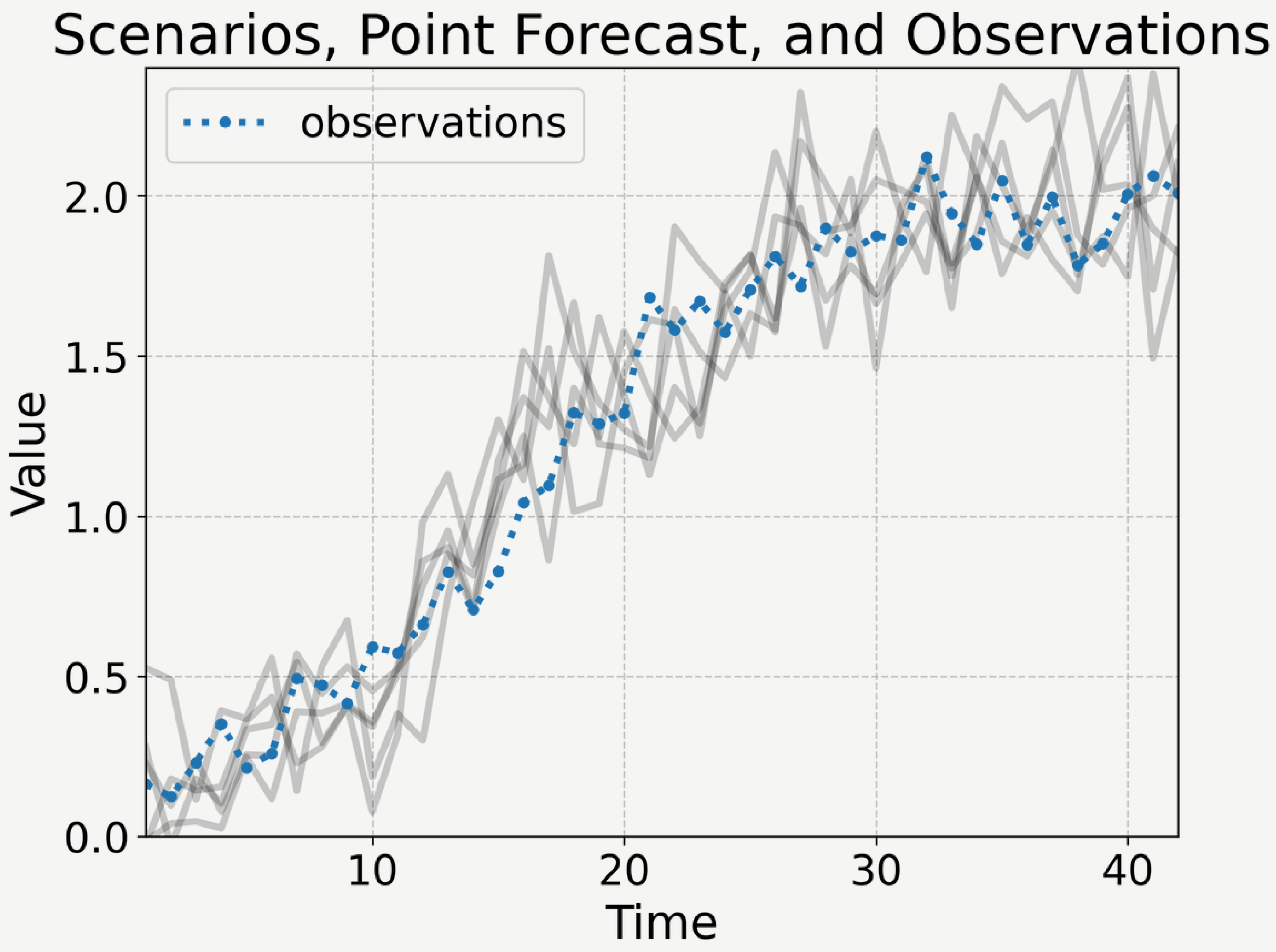
where:

$\hat{\boldsymbol{\Psi}}$ : estimated multivariate predictive distribution of $\mathbf{X}_t$

$\boldsymbol{\xi}^{(i)}$ : the $i$-th trajectory (scenario)

$\mathbf{X}_t$ : vector of future observations

$\sim$ : denotes "sampled from"



# Methods for (Probabilistic) Forecasting



**Complexity** →

## Statistical

Time-tested, interpretable models with strong theoretical foundations

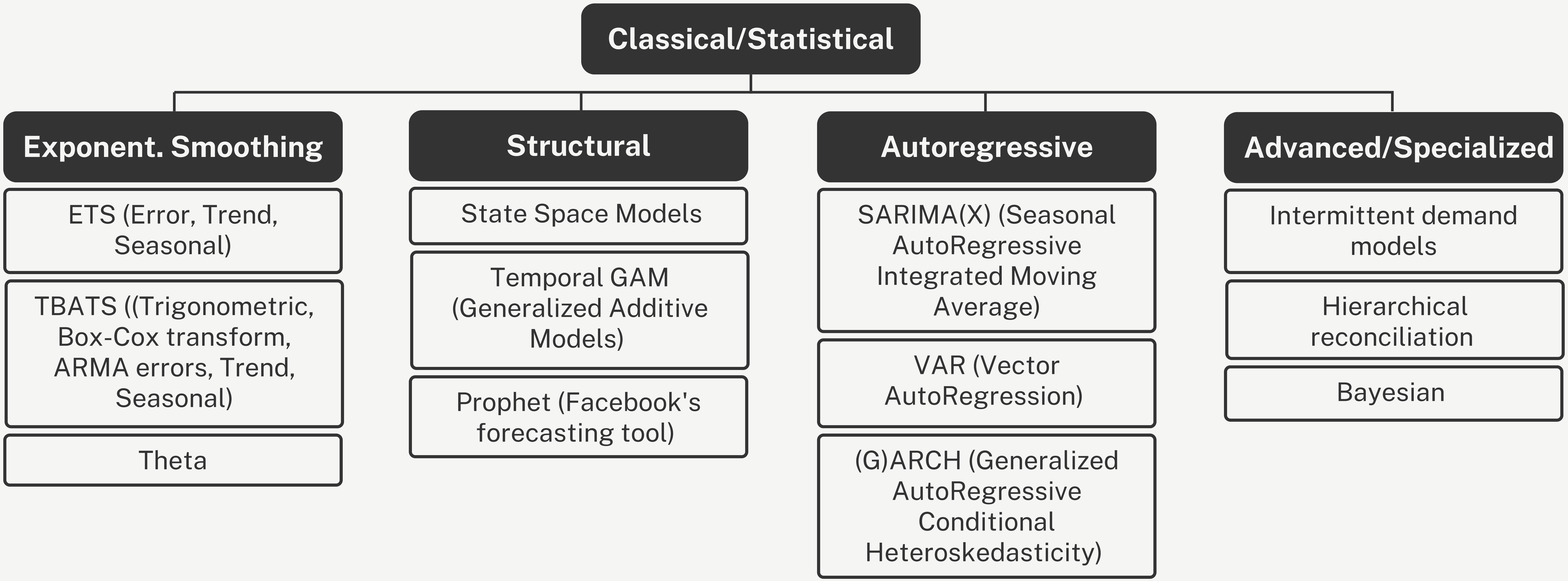
## ML

Flexible, data-driven approaches that can capture complex patterns

## DL

Powerful neural network-based models for handling large-scale, complex time series data

# Strategies for (Probabilistic) Forecasting



# Strategies for (Probabilistic) Forecasting

Forecasting for Social Good · Since 2020 ·

**Machine Learning**

- **Tree-Based**
  - Random Forests
  - Gradient Boosting Machines (XGBoost, LightGBM, CatBoost)
  - Decision Trees
- **Ensemble**
  - Bagging (eg. random forests)
  - Boosting (eg. AdaBoost, Gradient Boosting)
  - Stacking (combining predictions for models)
- **Reduction**
  - Time Series --> Supervised Learning
- **Other**
  - SVR (Support Vector Regression)
  - kNN (k-Nearest Neighbours)
  - Gaussian Processes

## Strategies for (Probabilistic) Forecasting



- **Deep learning**
  - **RNNs**
    - LSTM (Long Short-Term Memory)
    - GRU (Gated Recurrent Unit)
    - DeepAR (Amazon's RNN-based model)
  - **Transformers**
    - TFT (Temporal Fusion Transformer)
    - Informer, Autoformer, Pyraformer
    - (TST) Time Series Transformer
  - **LLMs**
    - Chronos, LagLlama, Moirai
    - Multimodal models: MomentFM, TimesFM
  - **Hybrid/Specialized**
    - Neural Prophet (Facebook's hybrid model)
    - NBEATS (Neural Basis Expansion Analysis)
    - ESRNN (Exponential Smoothing RNN)

# *Evaluating Forecasts*

## Forecasting Performance Metrics



| Point | Probabilistic |
| --- | --- |
| $\text{bias}(k) = \frac{1}{T}\sum_{t=1}^{T} e_{t+k\|t}$ | $L_\alpha(y,q) = \begin{cases} \alpha(y-q) & \text{if } y \geq q \\ (1-\alpha)(q-y) & \text{if } y < q \end{cases}$ |
| $\text{MAE}(k) = \frac{1}{T}\sum_{t=1}^{T} \|e_{t+k\|t}\|$ | $\text{CRPS} = \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty}(F_t(y) - \mathbf{1}\{y \geq y_t\})^2 dy$ |
| $\text{RMSE}(k) = \left[\frac{1}{T}\sum_{t=1}^{T} e_{t+k\|t}^2\right]^{1/2}$ | $\text{LogS} = -\frac{1}{T}\sum_{t=1}^{T}\log(p_t(y_t))$ |

## Key attributes of high-quality probabilistic forecasts

- **Reliability (Calibration)**
  - Consistency between forecasted probabilities and observed frequencies

## Key attributes of high-quality probabilistic forecasts

- **Reliability (Calibration)**
  - Consistency between forecasted probabilities and observed frequencies

- **Sharpness (Precision)**
  - Concentration of the predictive distribution

## Key attributes of high-quality probabilistic forecasts

- **Reliability (Calibration)**
  - Consistency between forecasted probabilities and observed frequencies
- **Sharpness (Precision)**
  - Concentration of the predictive distribution
- **Skill (Overall Performance)**
  - Improvement over a reference forecast (e.g., climatology)

## Key attributes of high-quality probabilistic forecasts

- **Reliability (Calibration)**
  - Consistency between forecasted probabilities and observed frequencies
- **Sharpness (Precision)**
  - Concentration of the predictive distribution
- **Skill (Overall Performance)**
  - Improvement over a reference forecast (e.g., climatology)

*Balancing act: Achieving reliability while maximizing sharpness and skill*

## Traps to Avoid When Assessing Probabilistic Models

- **Oversimplifying Accuracy Metrics**
  - *Issue:* Reducing probabilistic forecasts to binary outcomes loses valuable information
  - *Example:* Merely counting hits within a confidence interval overlooks forecast precision

## Traps to Avoid When Assessing Probabilistic Models

Forecasting for Social Good · Since 2020 ·

- **Oversimplifying Accuracy Metrics**
  - *Issue:* Reducing probabilistic forecasts to binary outcomes loses valuable information
  - *Example:* Merely counting hits within a confidence interval overlooks forecast precision

- **Misapplying Traditional Error Measures**
  - *Issue:* Using point-forecast metrics on sampled probabilities can lead to biased assessments
  - *Technical note:* $E[f(X)] \neq f(E[X])$ due to Jensen's inequality

## Traps to Avoid When Assessing Probabilistic Models

- **Oversimplifying Accuracy Metrics**
  - *Issue:* Reducing probabilistic forecasts to binary outcomes loses valuable information
  - *Example:* Merely counting hits within a confidence interval overlooks forecast precision

- **Misapplying Traditional Error Measures**
  - *Issue:* Using point-forecast metrics on sampled probabilities can lead to biased assessments
  - *Technical note:* $E[f(X)] \neq f(E[X])$ due to Jensen's inequality

- **Overestimating Model Certainty**
  - *Issue:* Assuming the model's uncertainty estimates are infallible
  - *Caution:* Probabilistic outputs require critical evaluation, not blind acceptance

## Time Series Cross-Validation

| [XXXXX] [-----] [-----] [-----] [-----] | **Fold 1**

| [-----] [XXXXX] [-----] [-----] [-----] | **Fold 2**

| [-----] [-----] [XXXXX] [-----] [-----] | **Fold 3**

| [-----] [-----] [-----] [XXXXX] [-----] | **Fold 4**

| [-----] [-----] [-----] [-----] [XXXXX] | **Fold 5**

*K-Fold Cross Validation*

Forecasting for Social Good
Since 2020

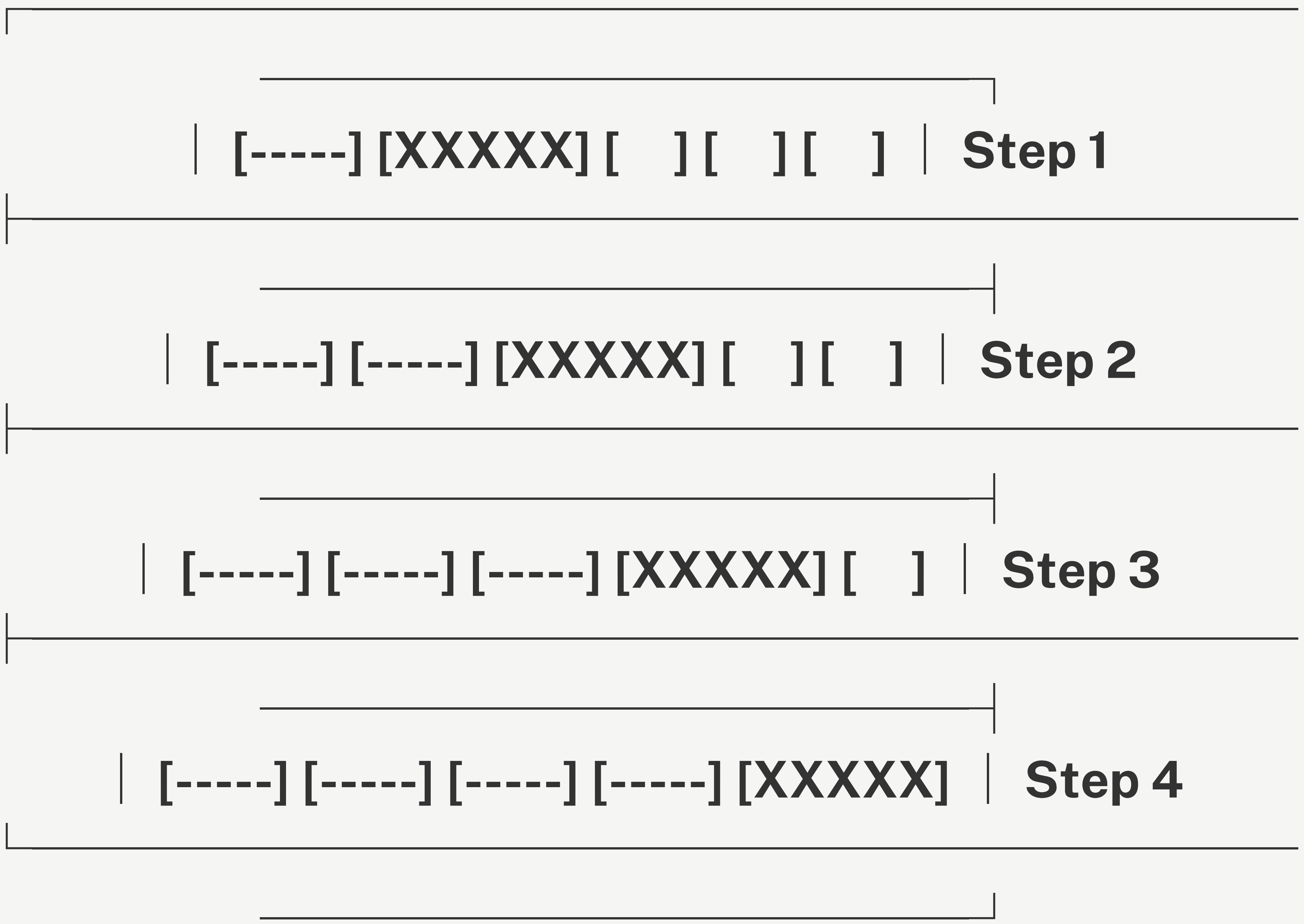
## Time Series Cross-Validation

```
[XXXXX] [-----] [-----] [-----] [-----]   Fold 1
[-----] [XXXXX] [-----] [-----] [-----]   Fold 2
[-----] [-----] [XXXXX] [-----] [-----]   Fold 3
[-----] [-----] [-----] [XXXXX] [-----]   Fold 4
[-----] [-----] [-----] [-----] [XXXXX]   Fold 5
```

*K-Fold Cross Validation*

## Time Series Cross-Validation

- Walk-Forward (Rolling Window) CV
- Expanding Window CV
- Gap Walk-Forward CV
- Purged CV
- Gap Walk-Forward with Purging CV
- Combinatorial Purged CV




*Walk-Forward Cross Validation*

# Python for TS

# A unified framework