Evaluating Uncertainty Requirement Performance

Amber Motley, Director Short Term Forecasting

Market Surveillance Committee Meeting

General Session

April 11, 2024

Outline

- Problem statement
- Hscore proposal
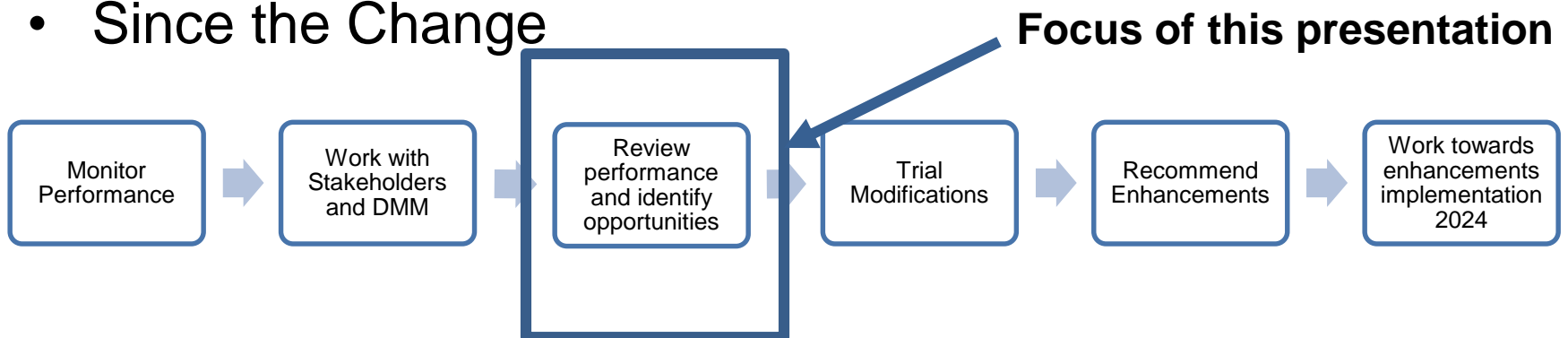- Illustrative use case

California ISO

Background

In February of 2023, per stakeholder request, the ISO enhanced the uncertainty modeling methodology to incorporate **weather forecast information** in lieu of history only when determining flexible product requirements (FRP).
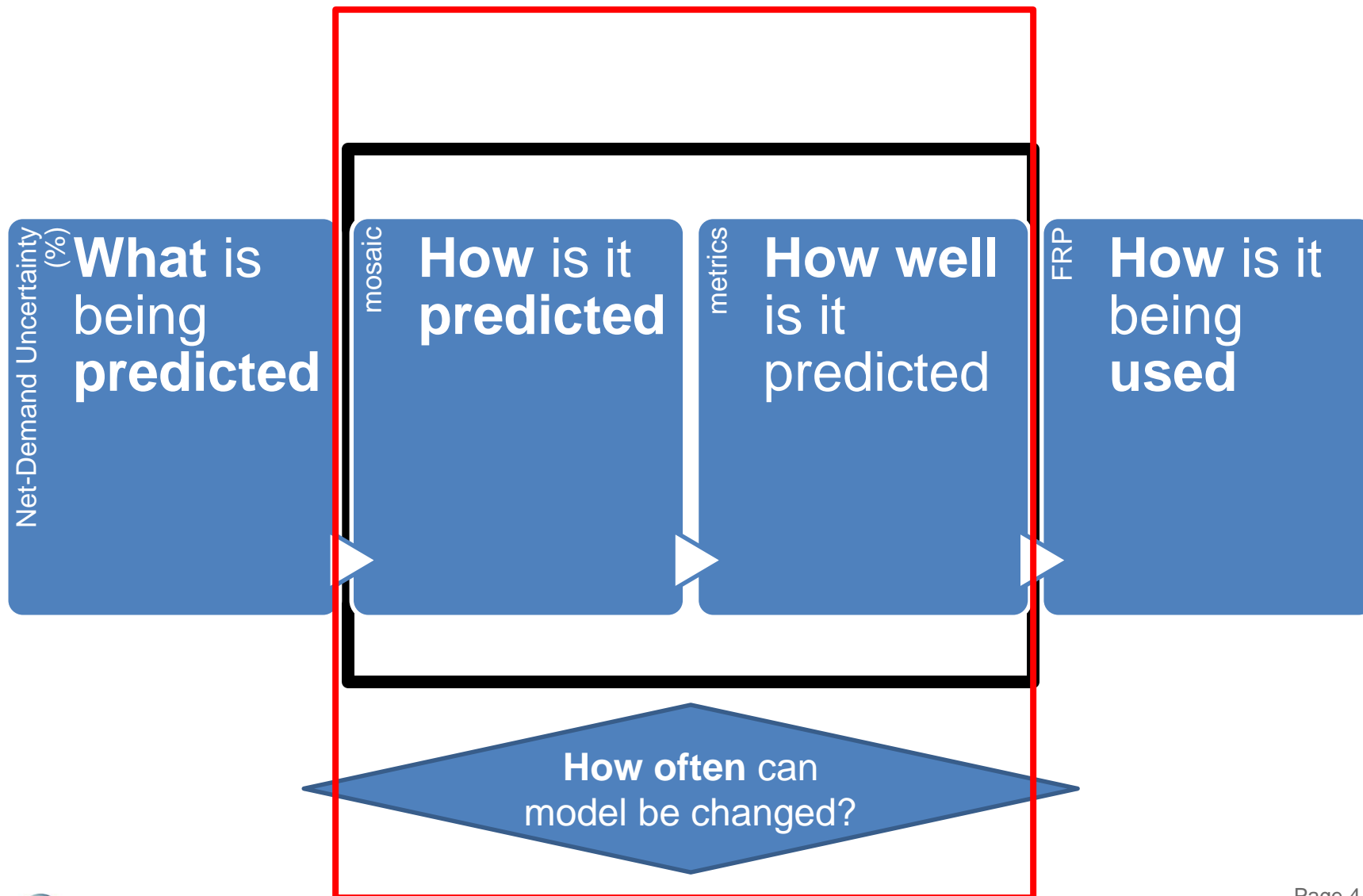
- Mosaic methodology incorporated into:

Real-Time (WEIM)	Day-Ahead
FRP Requirements	RUC Uncertainty
Resource Sufficiency Evaluation	Future Imbalance Reserve (IR) Requirements

- Since the Change

Focus of this presentation

Monitor Performance → Work with Stakeholders and DMM → Review performance and identify opportunities → Trial Modifications → Recommend Enhancements → Work towards enhancements implementation 2024

California ISO

ISO Public

Background

What is being **predicted**

Net-Demand Uncertainty (%)

How is it **predicted**

mosaic

How well is it predicted

metrics

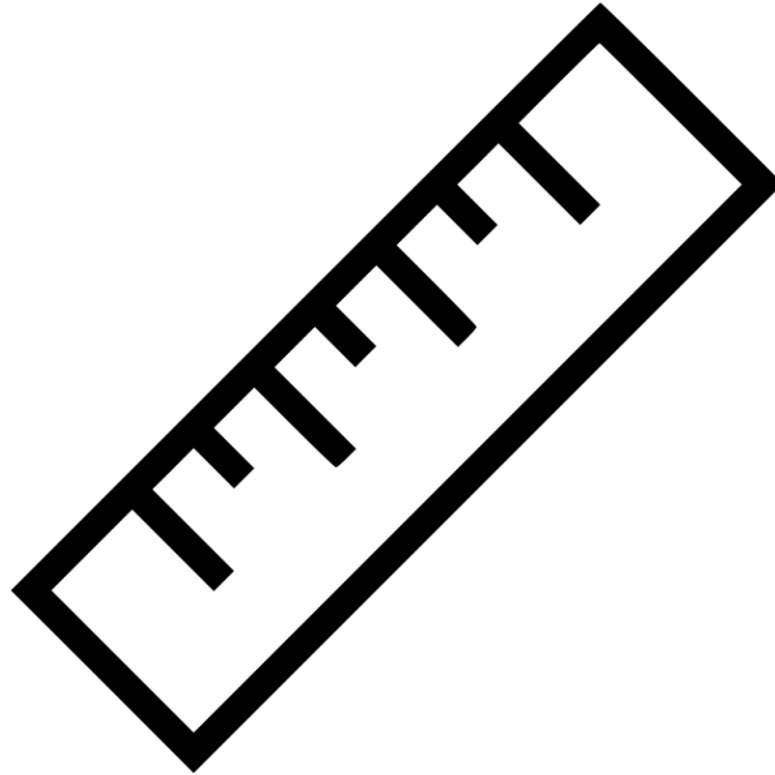**How** is it being **used**

FRP

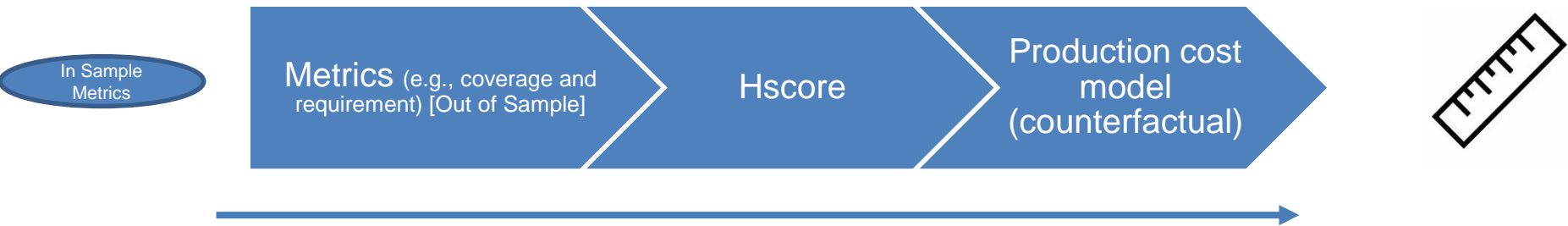**How often** can model be changed?

California ISO

How well is [the model] predicted?

Why is this **currently** a challenging task?

- We are broadly tasked w/ accurately predicting net load uncertainty while **minimizing cost and maintaining reliability**

- Comparing **this objective between alternative models is difficult …**
  - No single observable metric can capture this…
  - Further complicated when considering peak time, up vs. down, multiple BAAs, seasonality etc.

- Therefore, the ISO is working to enhance analysis to **definitively observe** and **communicate a better model**

- We propose an analog, for scoring performance of **alternative models**, called Hscore

California ISO

ISO Public

Observation and Communication of Performance

| In Sample Metrics | → | Metrics (e.g., coverage and requirement) [Out of Sample] | Hscore | Production cost model (counterfactual) |

Less difficult to construct — More difficult to construct

Less compute — More compute

Less information — More information

Less efficient to communicate — More efficient to communicate

California ISO

ISO Public

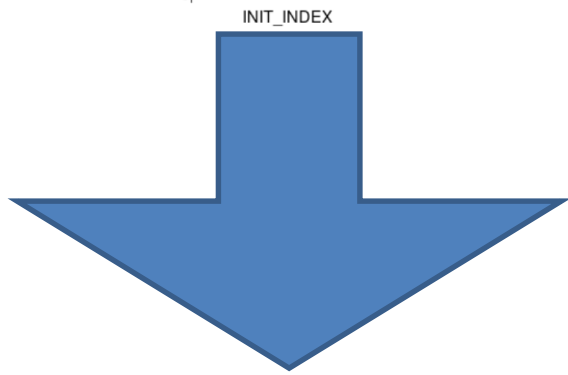# Hierarchical Scoring (Hscore) – Blueprint

- Decide all relevant **metrics**
- **Map** the metrics to **Calibration**, **Informativeness**, or **Cost**
- **Bin** model performance
  - Time of day (e.g., peak , ramp)
  - Directional consideration
  - Sample period length
  - BAA
- **Normalize** metrics
- **Weight** all metrics, robustly
- Generate **benchmark** forecasts
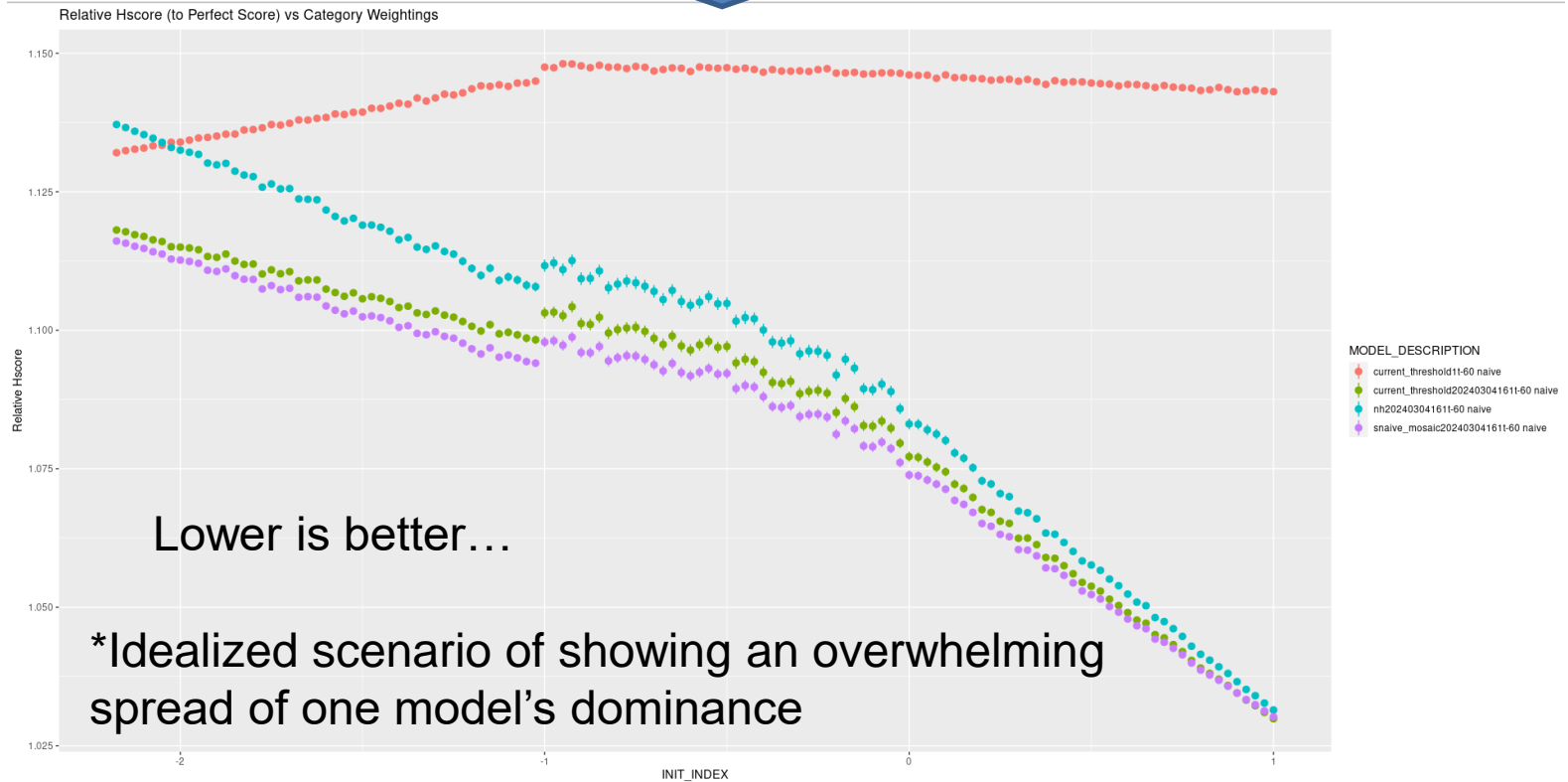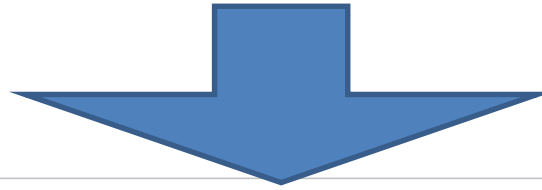 - Naïve, histogram and "perfect" model …

ISO Public

USE CASE

California ISO
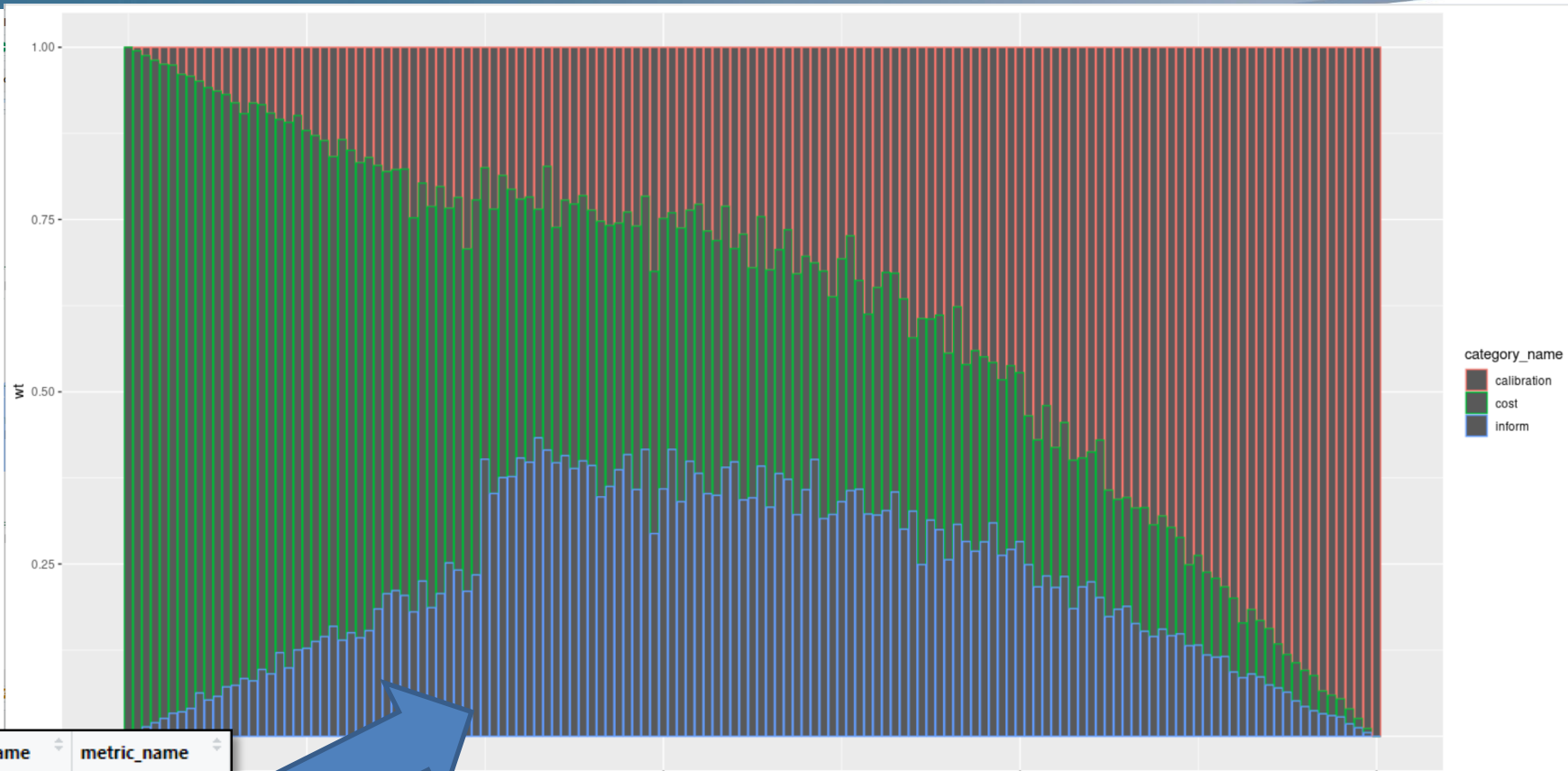
category_name	metric_name
calibration	cover
calibration	pb_bi
calibration	fuzzy_cover
cost	requirement
cost	exceed
inform	winkler
inform	breakpoint
inform	spearman
inform	pearson
usability	movement
usability	freq_threshold
usability	leadtime

ISO Public

Plot Relative Hscore by Different Candidate Weights

Relative Hscore (to Perfect Score) vs Category Weightings

Lower is better…

*Idealized scenario of showing an overwhelming spread of one model's dominance

California ISO

Conclusion

- Through this methodology we can robustly **observe** and **communicate** a model being superior against alternative models

    - Additionally, mosaic consistently outperforms histogram and naïve benchmarks

California ISO

ISO Public