Okay, here's the content formatted for a Google Slides presentation. You can copy and paste the text within each section into a new slide in Google Slides.

Slide 1: Title Slide

- Title: Evaluating Univariate Forecasts: A Comprehensive Metric Guide
- Subtitle: Bias, Data Anomaly Detection, Residual Anomaly Detection, Direction Accuracy (MDA), Continuous Ranked Probability Score (CRPS), Sliding Jensen-Shannon Divergence (JSD)

Slide 2: Introduction - Beyond Basic Accuracy

- Title: The Need for a Comprehensive Evaluation
- Bullet Points:
    - Why Use Multiple Metrics?
    - Standard metrics (MAPE, RMSE) measure average error magnitude but miss crucial aspects of forecast quality.[1]
    - A comprehensive evaluation requires a "scorecard" approach, using diverse metrics to understand different facets of performance.[4]
    - This helps build trust, identify specific issues (bias, instability, poor uncertainty estimates), and guide targeted improvements.[6]
    - Focus Areas: Systematic errors, data quality, model fit, trend prediction, probabilistic accuracy, and data stability over time.[7]

Slide 3: Forecast Bias (1/2)

- Title: Forecast Bias
- Definition:
    - A consistent, systematic tendency to overestimate or underestimate actual outcomes.[7]
    - Indicates forecast errors are not random but follow a persistent directional pattern.[13]
    - Distinct from random errors that fluctuate around the actual value.[15]
- Purpose:
    - Identify Systematic Errors: Pinpoints persistent, directional flaws in the model or process.[7]
    - Inform Adjustments: Guides corrections (increase forecast if consistently low, decrease if consistently high).[7]
    - Improve Planning: Reduces risks of stockouts or excess inventory; enhances resource and budget accuracy.[14]
    - Assess Usability: A fundamental check – high bias makes forecasts unreliable.[7]

Slide 4: Forecast Bias (2/2)

- Title: Forecast Bias (Continued)
- Mathematical Computation:
    - Mean Error (ME): Average of (Forecast - Actual). Simple, direct measure.[14]
        - Formula: Bias=n1∑i=1n(Fi−Ai) [7]
    - Percentage Bias: Bias relative to actual values. Useful for comparing across different scales.[15]
    - Tracking Signal (TS): Monitors bias accumulation over time. Ratio of cumulative error to Mean Absolute Deviation (MAD).[7]
        - Formula: TSt=MADt∑i=1t(Fi−Ai)
- Uniqueness & Advantage:
    - Unique Focus: Direction and consistency of errors, not just magnitude.[16]
    - Main Advantage: Directly flags systematic problems needing intervention (model, data, human judgment).[12] Leads to more dependable planning inputs.[14]
- Team Actions:
    - Data Science: Investigate model structure/parameters; implement corrections.[13]
    - Data Engineering: Verify data pipelines and accuracy.[14]
    - Business Stakeholders: Identify human bias sources; apply informed adjustments; understand business impact.[12]

Slide 5: Data Anomaly Detection (1/2)

- Title: Data Anomaly Detection
- Definition:
    - Identifying data points or patterns in the input time series that deviate significantly from normal behavior *before* forecasting.[8]
- Types:
    - Point Anomalies: Single unusual values.[21]
    - Contextual Anomalies: Unusual for the specific time/context.[20]
    - Collective Anomalies: Unusual sequence patterns.[20]
- Purpose:
    - Data Quality Assurance: Ensures clean, representative input data for models; prevents "garbage-in, garbage-out".[23]
    - Identify Critical Events: Flags real-world events like fraud, system errors, security incidents, equipment failures.[8]
    - Uncover Opportunities/Shifts: Reveals unexpected market movements or impacts of external events.[8]

Slide 6: Data Anomaly Detection (2/2)

- Title: Data Anomaly Detection (Continued)
- Detection Methods Overview:
  - Statistical: Z-score, modified Z-score, percentile ranges.[21]
  - Decomposition: Identify outliers in the residual component of input data (e.g., using STL).[19]
  - Machine Learning (Unsupervised): Clustering, PCA, Autoencoders, Isolation Forest, One-Class SVM.[8]
  - Machine Learning (Supervised/Semi-supervised): Use labeled data (if available) or train on normal data only.[28]
  - Distance/Density-based: k-NN, LOF, DBSCAN.[25]
- Uniqueness & Advantage:
  - Unique Focus: Scrutinizes the raw input data itself, independent of the forecasting model.[21]
  - Main Advantage: Critical data quality filter; protects forecast integrity; surfaces real-world event signals.[21]
- Team Actions:
  - Data Engineering: Implement validation/cleaning; investigate root causes; monitor ingestion.[21]
  - Data Science: Define handling strategy (remove, impute, flag); evaluate impact on models; select/tune algorithms.[21]
  - Business Stakeholders: Provide context for interpretation; decide if error or genuine event.[20]

Slide 7: Residual Anomaly Detection (1/2)

- Title: Residual Anomaly Detection
- Definition:
  - Identifying instances where the forecast error (Residual = Actual - Forecast) is unexpectedly large or erratic.[9]
  - Focuses on the part of the data the model failed to explain, unlike data anomaly detection which looks at inputs.[31]
  - Indicates a significant mismatch between prediction and reality, even with normal input data.[32]
- Purpose:
  - Assess Model Fit: Highlights where the model performs poorly.[32]
  - Identify Unforeseen Events: Signals unexpected shocks or structural breaks not captured by the model.[33]
  - Flag Deviations from Baseline: Pinpoints observations diverging significantly from the model's learned expectation.[31]

    - Improve Model Robustness (Online Learning): Helps decide whether to exclude surprising observations from model updates.[33]

Slide 8: Residual Anomaly Detection (2/2)

- Title: Residual Anomaly Detection (Continued)
- Detection Methods:
    - Decomposition-based: Apply outlier tests (e.g., Tukey's fence) to the residual component from time series decomposition (e.g., STL).[31]
    - Model-Specific Residual Analysis: Analyze residuals from models (ARIMA, ETS). Flag points outside statistical bounds (e.g., ±3 std dev, MAD-based) assuming residuals should be white noise.[24]
    - Forecast Distribution Comparison: For probabilistic forecasts, flag actuals falling in the extreme tails (low probability regions) of the predicted distribution.[33]
- Uniqueness & Advantage:
    - Unique Focus: Examines the unexplained variance *after* applying the model.[31]
    - Main Advantage: Provides direct feedback on model limitations and assumptions; guides model refinement or alerts to significant deviations.[32]
- Team Actions:
    - Data Science: Investigate causes of large residuals; refine model or detect structural breaks; decide handling strategy for updates.[32]
    - Business Stakeholders: Provide context explaining large errors.[32]
    - Data Engineering: Ensure accurate alignment of forecast and actuals data for residual calculation.[31]

Slide 9: Direction Accuracy (MDA) (1/2)

- Title: Direction Accuracy (MDA)
- Definition:
    - Mean Directional Accuracy (MDA) measures how accurately a forecast predicts the direction of change (increase or decrease) compared to the previous period.[5]
    - Focuses solely on the correctness of the predicted trend direction, ignoring the magnitude of change.[5]
- Purpose:
    - Assess Trend Capture: Determines if the model successfully captures up/down movements.[5]

    - Evaluate Directional Usefulness: Crucial when the direction of change is key for decisions (e.g., economics, finance, inventory planning).[5]
    - Benchmark Directional Prediction: Assesses if the forecast offers directional information beyond random chance (MDA > 0.5).[35]

Slide 10: Direction Accuracy (MDA) (2/2)

- Title: Direction Accuracy (MDA) (Continued)
- Mathematical Computation:
    - Compares the sign of change in actuals vs. the sign of change in forecasts.[36]
        - Formula: $\text{MDA} = \frac{1}{N-1} \sum_{t=2}^{N} \mathbf{1}[\text{sign}(A_t - A_{t-1}) == \text{sign}(F_t - F_{t-1})]$
    - Alternative: Compare sign of actual change vs. sign of (Current Forecast - Previous Actual).[34]
        - Formula: $\text{MDA} = \frac{1}{N} \sum_{t} \mathbf{1}[\text{sign}(A_t - A_{t-1}) == \text{sign}(F_t - A_{t-1})]$
    - Result: 0 to 1 (or 0% to 100%). 0.5 = random guess.[35]
- Uniqueness & Advantage:
    - Unique Focus: Exclusively measures direction of change, ignoring error magnitude.[5]
    - Main Advantage: Unambiguous measure of trend prediction ability, critical for strategic decisions often missed by magnitude metrics.[5]
- Team Actions:
    - Data Science: Use alongside magnitude metrics; investigate models with poor MDA.[36]
    - Business Stakeholders: Gauge confidence for directional decisions; compare MDA vs. 0.5 benchmark.[34]
    - Data Engineering: Ensure correct data alignment for difference calculations.[36]

Slide 11: Continuous Ranked Probability Score (CRPS) (1/2)

- Title: Continuous Ranked Probability Score (CRPS)
- Definition:
    - Evaluates the accuracy of a probabilistic forecast (which provides a full probability distribution) against a single, realized scalar observation.[37]
    - Measures the "distance" between the predicted Cumulative Distribution Function (CDF) and the actual outcome's step-function CDF.[38]
- Purpose:

    - Evaluate Probabilistic Forecast Quality: Assesses both location (central tendency) and spread (uncertainty) of the predicted distribution.[10]
    - Compare Probabilistic Models: Ranks the performance of different models providing distributional forecasts.[40]
    - Provide Summary Score: Offers a single score representing how well the entire predicted distribution matched reality.[42]

Slide 12: Continuous Ranked Probability Score (CRPS) (2/2)

- Title: Continuous Ranked Probability Score (CRPS) (Continued)
- Mathematical Computation:
    - Integral of the squared difference between forecast CDF $F(u)$ and observation's step-function CDF $H(u-y)$ [40]:
        - Formula: $\mathrm{CRPS}(F,y)=\int_{-\infty}^{\infty}[F(u)-H(u-y)]^2du$
    - Alternative form (Expected Absolute Error) [43]:
        - Formula: $\mathrm{CRPS}(F,y)=E_F[|X-y|]-\frac{1}{2}E_F[|X-X^*|]$
    - Units: Same as the forecast variable.[43] Lower is better (0 = perfect).[44]
    - Generalizes MAE: Reduces to MAE for deterministic (point) forecasts.[43]
- Uniqueness & Advantage:
    - Unique Focus: Evaluates the entire predictive probability distribution, not just a point or quantile.[10]
    - Main Advantage: It's a proper scoring rule, encouraging honest probability reporting.[44] Bridges point and probabilistic forecast evaluation.[10] Rewards accurate and well-calibrated uncertainty estimates.[37]
- Team Actions:
    - Data Science: Use as primary metric for probabilistic models; analyze scores; potentially decompose CRPS for deeper insights.[42]
    - Business Stakeholders: Understand it assesses the predicted range vs. actual outcome; build confidence in uncertainty estimates.[10]
    - Data Engineering: Ensure pipelines handle distributional forecast outputs.[39]

Slide 13: Sliding Jensen-Shannon Divergence (JSD) (1/2)

- Title: Sliding Jensen-Shannon Divergence (JSD)
- Definition:
    - Jensen-Shannon Divergence (JSD): Measures the similarity between two probability distributions.[46] Symmetric and bounded (0 to 1 or log(2)).[47] 0 = identical distributions.[48]

    - Sliding JSD: Applies JSD over time using a sliding window.[50] Compares the data distribution in a current window to a previous or reference window.[11]
- Purpose:
    - Detect Distribution Drift / Concept Drift: Identifies changes over time in input data statistics (covariate drift) or input-output relationships (concept drift, often via residuals).[51] Crucial as models assume stable data.[53]
    - Monitor Data Stability: Assesses if recent data characteristics match the training baseline.[53]
    - Identify Anomalous Regimes: Detects periods where overall data behavior (distribution shape) changes significantly.[55]
    - Trigger Model Adaptation: Signals potential need for model retraining or replacement due to environmental changes.[57]

Slide 14: Sliding Jensen-Shannon Divergence (JSD) (2/2)

- Title: Sliding Jensen-Shannon Divergence (JSD) (Continued)
- Mathematical Computation:
    - Core JSD: Based on KL Divergence. $\text{JSD}(P || Q) = \frac{1}{2} \text{KL}(P || M) + \frac{1}{2} \text{KL}(Q || M)$, where $M=0.5\times(P+Q)$.[48]
    - Time Series Application: Requires deriving distributions from time windows (e.g., histograms/binning [58], symbolic mapping [50], sketches [11]).
    - Sliding Window: Calculate JSD between current and reference window distributions at each step.[50]
    - Interpretation: Monitor the JSD time series. Spikes or sustained increases signal distribution divergence, often compared to a threshold.[55]
- Uniqueness & Advantage:
    - Unique Focus: Quantifies difference between probability distributions over time; sensitive to changes in shape (variance, skew, modality).[61]
    - Main Advantage: Directly addresses non-stationarity/drift.[62] Robust and interpretable compared to raw KL divergence.[46] Principled way to detect complex changes.[58]
- Team Actions:
    - Data Science: Monitor inputs (covariate drift) and residuals (concept drift); use alerts to trigger investigation/retraining; evaluate distribution estimation methods.[64]
    - Data Engineering: Build pipelines for JSD calculation; monitor upstream data sources for changes.[55]
    - Business Stakeholders: Understand drift alerts signal environmental change; discuss implications and responses.[65]

Slide 15: Recap - Why Each Metric is Unique

- Title: Understanding Metric Uniqueness
- Bullet Points:
  - Bias: Focuses only on the consistent direction of errors (over/under). *Are we systematically wrong in one direction?* [16]
  - Data Anomaly: Focuses only on the input data quality *before* forecasting. *Is the history we're using reliable?* [21]
  - Residual Anomaly: Focuses only on forecast errors *after* modeling. *Where did the model fail significantly?* [31]
  - Direction Accuracy (MDA): Focuses only on the direction of change (up/down). *Did we predict the trend correctly, regardless of magnitude?* [5]
  - CRPS: Focuses only on the entire probabilistic distribution. *How well did our predicted range of possibilities match the single outcome?* [10]
  - Sliding JSD: Focuses only on changes in the data's distribution shape over time. *Is the underlying process stable or drifting?* [11]

Slide 16: Recap - Advantages & Team Actions

- Title: Metric Scorecard: Advantages and Team Responsibilities
- Table:

| Metric | Main Advantage | Data Science Actions | Data Engineering Actions | Business Stakeholder Actions |
|---|---|---|---|---|
| Bias | Flags systematic flaws needing intervention [12] | Adjust model/parameters [13] | Verify data pipelines [14] | Identify human bias, understand impact [12] |
| Data Anomaly | Protects forecast integrity, surfaces events [21] | Define handling strategy, select algorithms [21] | Implement cleaning, investigate sources [21] | Provide context, interpret findings [20] |

| | | | | |
|---|---|---|---|---|
| Residual Anomaly | Direct feedback on model limits/failures [32] | Refine model, detect breaks, manage updates [32] | Ensure data alignment [31] | Provide context for errors [32] |
| MDA | Clear measure of trend prediction ability [5] | Evaluate trend capture [36] | Ensure data for diffs [36] | Assess directional confidence [34] |
| CRPS | Proper scoring rule for probabilistic forecasts [44] | Evaluate/tune probabilistic models [42] | Handle distributional outputs [39] | Build confidence in uncertainty estimates [10] |
| Sliding JSD | Detects non-stationarity/distribution drift [62] | Monitor drift, trigger investigation/retraining [53] | Build JSD pipelines, monitor sources [11] | Understand environmental shifts, discuss response [57] |