Statistical Metrics Help Document

1. Mean Squared Error (MSE)

Formula: $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2$

Meaning: MSE measures the average squared difference between two time-dependent datasets x and y. Smaller MSE values indicate better alignment between the datasets.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [49.8, 50.0, 49.9] MPa.
  Dataset 2 (predicted): [50.0, 49.9, 49.8] MPa.

  Step 1: Calculate squared differences:
  $(50.0 - 49.8)^2 = 0.04$
  $(49.9 - 50.0)^2 = 0.01$
  $(49.8 - 49.9)^2 = 0.01$

  Step 2: Compute MSE*:*
  $\mathrm{MSE} = \frac{0.04+0.01+0.01}{3} = 0.02$ MPa².

  Step 3: Interpret Result:
  An MSE of 0.02 MPa² suggests a small difference between the measured and predicted pressure datasets.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [890, 910, 900] °C.
  Dataset 2 (predicted): [900, 900, 905] °C.

  Step 1: Calculate squared differences:
  $(900 - 890)^2 = 100$
  $(900 - 910)^2 = 100$
  $(905 - 900)^2 = 25$

  Step 2: Compute MSE*:*
  $\mathrm{MSE} = \frac{100+100+25}{3} = 75$ °C².

  Step 3: Interpret Result:
  An MSE of 75 °C² indicates moderate discrepancies between the datasets, warranting calibration adjustments.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1980, 2020, 2005] microstrain.
  Dataset 2 (predicted): [2000, 2010, 2000] microstrain.

  Step 1: Calculate squared differences:
  $(2000 - 1980)^2 = 400$
  $(2010 - 2020)^2 = 100$
  $(2000 - 2005)^2 = 25$

  Step 2: Compute MSE*:*
  $\mathrm{MSE} = \frac{400+100+25}{3} = 175$ microstrain².

  Step 3: Interpret Result:
  An MSE of 175 microstrain² shows acceptable alignment but suggests minor model improvements.

2. Root Mean Squared Error (RMSE)

Formula: $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$

Meaning: RMSE provides error magnitude in the same units as the data, allowing easier interpretation of how well two time-dependent datasets align.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [49.8, 50.0, 49.9] MPa.
  Dataset 2 (predicted): [50.0, 49.9, 49.8] MPa.

  Step 1: Compute MSE:
  Using the $\mathrm{MSE} = 0.02$ MPa² calculated in the MSE example.

  Step 2: Compute RMSE:
  $\mathrm{RMSE} = \sqrt{0.02} = 0.141$ MPa.

  Step 3: Interpret Result:
  The RMSE of 0.141 MPa indicates minor differences between the measured and predicted datasets, validating the accuracy of the compressor model.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [890, 910, 900] °C.
  Dataset 2 (predicted): [900, 900, 905] °C.

  Step 1: Compute MSE:
  Using the $\mathrm{MSE} = 75$ °C² calculated in the MSE example.

  Step 2: Compute RMSE:
  $\mathrm{RMSE} = \sqrt{75} = 8.66$ °C.

  Step 3: Interpret Result:
  The RMSE of 8.66 °C indicates acceptable deviations between datasets, but highlights a need for slight calibration adjustments in temperature predictions.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1980, 2020, 2005] microstrain.
  Dataset 2 (predicted): [2000, 2010, 2000] microstrain.

  Step 1: Compute MSE:
  Using the $\mathrm{MSE} = 175$ microstrain² calculated in the MSE example.

  Step 2: Compute RMSE:
  $\mathrm{RMSE} = \sqrt{175} \approx 13.23$ microstrain.

  Step 3: Interpret Result:
  The RMSE of 13.23 microstrain suggests acceptable alignment between datasets, but indicates room for improvement in the strain prediction model for better accuracy during high-stress tests.

3. Absolute Error

Formula: $\text{Absolute Error} = |x - y|$

Meaning: Absolute Error measures the exact deviation between two time-dependent datasets without considering the direction of the error.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [49.8, 50.0, 49.9] MPa.

Dataset 2 (predicted): [50.0, 49.9, 49.8] MPa.

Step 1: Calculate absolute errors:
$|50.0 - 49.8| = 0.2$ MPa
$|49.9 - 50.0| = 0.1$ MPa
$|49.8 - 49.9| = 0.1$ MPa

Step 2: Interpret Results:
The errors indicate minor deviations (0.1–0.2 MPa) between datasets, validating the compressor model for consistent pressure predictions.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [890, 910, 900] °C.
  Dataset 2 (predicted): [900, 900, 905] °C.

  Step 1: Calculate absolute errors:
  $|900 - 890| = 10$ °C
  $|900 - 910| = 10$ °C
  $|905 - 900| = 5$ °C

  Step 2: Interpret Results:
  Absolute errors of 5–10 °C suggest moderate alignment between datasets, highlighting the need for slight calibration improvements in the temperature model.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1980, 2020, 2005] microstrain.
  Dataset 2 (predicted): [2000, 2010, 2000] microstrain.

  Step 1: Calculate absolute errors:
  $|2000 - 1980| = 20$ microstrain
  $|2010 - 2020| = 10$ microstrain
  $|2000 - 2005| = 5$ microstrain

  Step 2: Interpret Results:
  Absolute errors of 5–20 microstrain indicate acceptable alignment, but emphasize the need for refining strain prediction models for high-stress environments.

4. Percentage Error

Formula: $\text{Percentage Error} = \frac{|x-y|}{x} \times 100\%$

Meaning: Percentage Error quantifies the relative error between two datasets as a percentage of the actual values, making it easier to compare across different scales.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [49.8, 50.0, 49.9] MPa.
  Dataset 2 (predicted): [50.0, 49.9, 49.8] MPa.

  Step 1: Calculate percentage errors:
  $\frac{|50.0-49.8|}{50.0} \times 100 = 0.4\%$
  $\frac{|49.9-50.0|}{50.0} \times 100 = 0.2\%$
  $\frac{|49.8-49.9|}{49.8} \times 100 = 0.2\%$

  Step 2: Interpret Results:
  Percentage errors of 0.2–0.4% indicate highly accurate pressure predictions, validating the compressor model's reliability under varying loads.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [890, 910, 900] °C.
  Dataset 2 (predicted): [900, 900, 905] °C.

  Step 1: Calculate percentage errors:
  $\frac{|900-890|}{900} \times 100 = 1.1\%$
  $\frac{|900-910|}{900} \times 100 = 1.1\%$
  $\frac{|905-900|}{900} \times 100 = 0.6\%$

  Step 2: Interpret Results:
  Errors of 0.6–1.1% suggest moderate alignment, but calibration refinements may improve the turbine inlet temperature model.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1980, 2020, 2005] microstrain.
  Dataset 2 (predicted): [2000, 2010, 2000] microstrain.

  Step 1: Calculate percentage errors:
  $\frac{|2000-1980|}{2000} \times 100 = 1.0\%$
  $\frac{|2010-2020|}{2010} \times 100 = 0.5\%$
  $\frac{|2000-2005|}{2000} \times 100 = 0.25\%$

  Step 2: Interpret Results:
  Errors of 0.25–1.0% highlight acceptable alignment but suggest slight refinements in the strain prediction model for high-stress blade operations.

5. Symmetric Mean Absolute Percentage Error (SMAPE)

Formula: $\text{SMAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|x_i - y_i|}{(|x_i| + |y_i|)/2} \times 100\%$

Meaning: SMAPE is particularly useful for comparing datasets with near-zero values, where percentage error can produce excessively high or undefined results. SMAPE normalizes the error symmetrically, providing balanced and interpretable results.

Practical Use Cases:

- **Compressor Discharge Pressure (Transient Startup):**
  Dataset 1 (measured): [1×10^{-6}, 2.0, 3.0, 3.5, 4.0, 4.5, 5.0, 5.2, 5.3, 5.4] MPa.
  Dataset 2 (predicted): [5×10^{-6}, 2.1, 2.9, 3.6, 3.8, 4.4, 5.1, 5.1, 5.4, 5.5] MPa.

  Step 1: Calculate SMAPE for each data point:
  $\frac{|5\times10^{-6} - 1\times10^{-6}|}{((5\times10^{-6}) + (1\times10^{-6}))/2} \times 100 = 133.33\%$
  $\frac{|2.1-2.0|}{(2.1+2.0)/2} \times 100 = 4.88\%$
  $\frac{|2.9-3.0|}{(2.9+3.0)/2} \times 100 = 3.37\%$
  $\frac{|3.6-3.5|}{(3.6+3.5)/2} \times 100 = 2.82\%$
  $\frac{|3.8-4.0|}{(3.8+4.0)/2} \times 100 = 5.13\%$
  $\frac{|4.4-4.5|}{(4.4+4.5)/2} \times 100 = 2.27\%$
  $\frac{|5.1-5.0|}{(5.1+5.0)/2} \times 100 = 1.96\%$
  $\frac{|5.1-5.2|}{(5.1+5.2)/2} \times 100 = 1.92\%$
  $\frac{|5.4-5.3|}{(5.4+5.3)/2} \times 100 = 1.88\%$
  $\frac{|5.5-5.4|}{(5.5+5.4)/2} \times 100 = 1.85\%$

  Step 2: Compute average SMAPE:
  $\text{SMAPE} = \frac{133.33+4.88+3.37+2.82+5.13+2.27+1.96+1.92+1.88+1.85}{10} = 15.74\%.$

  Step 3: Compare with Percentage Error:
  The first percentage error would yield: $\frac{|5\times10^{-6} - 1\times10^{-6}|}{1\times10^{-6}} \times 100 = 400.0\%$. The subsequent errors are relatively small, but the average percentage error would be heavily skewed by the first value. SMAPE greatly reduces the effect of this distortion.

Step 4: Interpret Results:
The SMAPE of 15.74% reflects the overall alignment of datasets during the startup phase, avoiding excessive emphasis on near-zero values.

- **Gas Turbine Inlet Temperature (Startup Phase):**
Dataset 1 (measured): [1×10^{-6}, 850, 855, 860, 870, 880, 885, 890, 895, 900] °C.
Dataset 2 (predicted): [3×10^{-6}, 845, 860, 855, 875, 878, 888, 889, 892, 905] °C.

Step 1: Calculate SMAPE for each data point:
$\frac{|3\times10^{-6}-1\times10^{-6}|}{((3\times10^{-6})+(1\times10^{-6}))/2} \times 100 = 133.33\%$
$\frac{|845-850|}{(845+850)/2} \times 100 = 0.59\%$
$\frac{|860-855|}{(860+855)/2} \times 100 = 0.58\%$
$\frac{|855-860|}{(855+860)/2} \times 100 = 0.58\%$
$\frac{|875-870|}{(875+870)/2} \times 100 = 0.57\%$
$\frac{|878-880|}{(878+880)/2} \times 100 = 0.45\%$
$\frac{|888-885|}{(888+885)/2} \times 100 = 0.34\%$
$\frac{|889-890|}{(889+890)/2} \times 100 = 0.11\%$
$\frac{|892-895|}{(892+895)/2} \times 100 = 0.17\%$
$\frac{|905-900|}{(905+900)/2} \times 100 = 0.28\%$

Step 2: Compute average SMAPE:
$\text{SMAPE} = \frac{133.33+0.59+0.58+0.58+0.57+0.45+0.34+0.11+0.17+0.28}{10} = 13.44\%.$

Step 3: Compare with Percentage Error:
Percentage error for the first data point is 200.0%. SMAPE mitigates this by balancing discrepancies across all points.

Step 4: Interpret Results:
The SMAPE of 13.44% demonstrates strong alignment during the cooldown phase while handling the near-zero values effectively.

- **Blade Strain Testing (Ramp-Up Loads):**
Dataset 1 (measured): [1×10^{-6}, 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000] microstrain.
Dataset 2 (predicted): [3×10^{-6}, 120, 480, 950, 1480, 1980, 2480, 2980, 3480, 3980] microstrain.

Step 1: Calculate SMAPE for each data point:
$\frac{|3\times10^{-6}-1\times10^{-6}|}{((3\times10^{-6})+(1\times10^{-6}))/2} \times 100 = 66.67\%$
$\frac{|120-100|}{(120+100)/2} \times 100 = 18.18\%$
$\frac{|480-500|}{(480+500)/2} \times 100 = 4.08\%$
$\frac{|950-1000|}{(950+1000)/2} \times 100 = 2.56\%$
$\frac{|1480-1500|}{(1480+1500)/2} \times 100 = 1.33\%$
$\frac{|1980-2000|}{(1980+2000)/2} \times 100 = 1.00\%$
$\frac{|2480-2500|}{(2480+2500)/2} \times 100 = 0.80\%$
$\frac{|2980-3000|}{(2980+3000)/2} \times 100 = 0.67\%$
$\frac{|3480-3500|}{(3480+3500)/2} \times 100 = 0.57\%$
$\frac{|3980-4000|}{(3980+4000)/2} \times 100 = 0.50\%$

Step 2: Compute average SMAPE:
$\text{SMAPE} = \frac{66.67+18.18+4.08+2.56+1.33+1.00+0.80+0.67+0.57+0.50}{10} = 9.64\%.$

Step 3: Compare with Percentage Error:
Percentage errors would yield: $\frac{|3\times10^{-6}-1\times10^{-6}|}{1\times10^{-6}} \times 100 = 200.0\%$, $\frac{|120-100|}{100} \times 100 = 20.0\%$, $\frac{|480-500|}{500} \times 100 = 4.0\%$, and smaller errors for subsequent data points. The average percentage error would be significantly higher than SMAPE due to the first value.

Step 4: Interpret Results:
The SMAPE of 9.64% reflects the alignment between measured and predicted datasets during ramp-up loading, avoiding the excessive influence of the near-zero initial load.

6. Weighted Mean Absolute Percentage Error (WMAPE)

Formula: $\text{WMAPE} = \frac{\sum_{i=1}^{n}|x_i-y_i|}{\sum_{i=1}^{n}x_i} \times 100\%$

Meaning: WMAPE quantifies the relative error between two datasets while accounting for the magnitude of the actual values. By weighting errors proportionally to the scale of the actual values, WMAPE avoids distortions caused by small values dominating the calculation.

Practical Use Cases:

- **Compressor Discharge Pressure (Startup Phase):**
Dataset 1 (measured): [1×10^{-6}, 2.0, 3.0, 3.5, 4.0, 4.5, 5.0, 5.2, 5.3, 5.4] MPa.
Dataset 2 (predicted): [5×10^{-6}, 2.1, 2.9, 3.6, 3.8, 4.4, 5.1, 5.1, 5.4, 5.5] MPa.

Step 1: Calculate weighted absolute errors:
$|5 \times 10^{-6} - 1 \times 10^{-6}| = 4 \times 10^{-6}$
$|2.1 - 2.0| = 0.1$
$|2.9 - 3.0| = 0.1$
$|3.6 - 3.5| = 0.1$
$|3.8 - 4.0| = 0.2$
$|4.4 - 4.5| = 0.1$
$|5.1 - 5.0| = 0.1$
$|5.1 - 5.2| = 0.1$
$|5.4 - 5.3| = 0.1$
$|5.5 - 5.4| = 0.1$

Step 2: Compute WMAPE:
$\text{WMAPE} = \frac{\sum|x_i-y_i|}{\sum x_i} \times 100 = \frac{4\times10^{-6}+0.1+0.1+0.1+0.2+0.1+0.1+0.1+0.1+0.1}{1\times10^{-6}+2.0+3.0+3.5+4.0+4.5+5.0+5.2+5.3+5.4} \times 100$
$= \frac{1.0}{39.9} \times 100 = 2.51\%.$

Step 3: Compare with Percentage Error:
The percentage error for the first data point is: $\frac{|5\times10^{-6}-1\times10^{-6}|}{1\times10^{-6}} \times 100 = 400.0\%$, which skews the overall average percentage error significantly. WMAPE avoids this distortion by weighting errors according to the total scale of the actual values.

Step 4: Interpret Results:
WMAPE of 2.51% reflects the overall alignment of datasets, avoiding disproportionate emphasis on the near-zero first value.

- **Gas Turbine Inlet Temperature (Startup Phase):**
Dataset 1 (measured): [1×10^{-6}, 850, 855, 860, 870, 880, 885, 890, 895, 900] °C.
Dataset 2 (predicted): [3×10^{-6}, 845, 860, 855, 875, 878, 888, 889, 892, 905] °C.

Step 1: Calculate weighted absolute errors:
$|3 \times 10^{-6} - 1 \times 10^{-6}| = 2 \times 10^{-6}$
$|845 - 850| = 5$
$|860 - 855| = 5$
$|855 - 860| = 5$
$|875 - 870| = 5$
$|878 - 880| = 2$
$|888 - 885| = 3$
$|889 - 890| = 1$
$|892 - 895| = 3$
$|905 - 900| = 5$

Step 2: Compute WMAPE:

$\mathrm{WMAPE} = \frac{\sum|x_i - y_i|}{\sum x_i} \times 100 = \frac{2\times10^{-6}+5+5+5+5+5+2+3+1+3+5}{1\times10^{-6}+850+855+860+870+880+885+890+895+900} \times 100$
$= \frac{39}{7890} \times 100 = 0.49\%.$

Step 3: Compare with Percentage Error:
The percentage error for the first data point is: $\frac{|3\times10^{-6}-1\times10^{-6}|}{1\times10^{-6}} \times 100 = 200.0\%$. WMAPE avoids this extreme value by weighting errors proportionally to the total scale of the dataset.

Step 4: Interpret Results:
WMAPE of 0.49% confirms strong alignment across the dataset, avoiding the distortion caused by near-zero values.

- **Blade Strain Testing (Ramp-Up Loads):**
  Dataset 1 (measured): [1×10^{-6}, 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000] microstrain.
  Dataset 2 (predicted): [2×10^{-6}, 95, 480, 950, 1480, 1980, 2480, 2980, 3480, 3980] microstrain.

  Step 1: Calculate weighted absolute errors:
  $|2 \times 10^{-6} - 1 \times 10^{-6}| = 1 \times 10^{-6}$
  $|95 - 100| = 5$
  $|480 - 500| = 20$
  $|950 - 1000| = 50$
  $|1480 - 1500| = 20$
  $|1980 - 2000| = 20$
  $|2480 - 2500| = 20$
  $|2980 - 3000| = 20$
  $|3480 - 3500| = 20$
  $|3980 - 4000| = 20$

  Step 2: Compute WMAPE:
  $\mathrm{WMAPE} = \frac{\sum|x_i - y_i|}{\sum x_i} \times 100 = \frac{1\times10^{-6}+5+20+50+20+20+20+20+20+20}{1\times10^{-6}+100+500+1000+1500+2000+2500+3000+3500+4000} \times 100$
  $= \frac{195}{14100} \times 100 = 1.38\%.$

  Step 3: Compare with Percentage Error:
  The percentage error for the first data point is: $\frac{|2\times10^{-6}-1\times10^{-6}|}{1\times10^{-6}} \times 100 = 100.0\%$. WMAPE avoids the disproportionate influence of this value, yielding a balanced result.

  Step 4: Interpret Results:
  WMAPE of 1.38% confirms the alignment between datasets during ramp-up loads, accounting for the magnitude of the actual values.

7. Pearson Correlation Coefficient (PCC)

Formula: $\mathrm{PCC} = \frac{\mathrm{Cov}(x,y)}{\sigma_x \cdot \sigma_y}$

Meaning: PCC quantifies the strength and direction of the linear relationship between two datasets. Values range from -1 (perfect negative correlation) to +1 (perfect positive correlation). A value of 0 indicates no linear correlation.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5] MPa.
  Dataset 2 (predicted): [2.8, 3.6, 4.2, 4.4, 5.1, 5.6, 3.0, 6.6, 7.1, 7.6] MPa.

  Step 1: Calculate means:
  Mean of Dataset 1 ($\bar{x}$): $\frac{3.0+3.5+\cdots+7.5}{10} = 5.25$
  Mean of Dataset 2 ($\bar{y}$): $\frac{2.8+3.6+\cdots+7.6}{10} = 5.20$

  Step 2: Calculate covariance ($\mathrm{Cov}(x,y)$):
  $\mathrm{Cov}(x,y) = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n} = 2.19$

  Step 3: Calculate standard deviations (σ_x and σ_y):
  $\sigma_x = 1.37, \sigma_y = 1.62$

  Step 4: Calculate PCC:
  $\mathrm{PCC} = \frac{\mathrm{Cov}(x,y)}{\sigma_x \cdot \sigma_y} = \frac{2.19}{1.37 \cdot 1.62} = 0.98$

  Step 5: Interpret Results:
  A PCC of 0.98 indicates a strong positive linear relationship, though some deviations (e.g., predicted value of 3.0 MPa for the seventh data point) suggest a need for further investigation in specific scenarios.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [850, 860, 870, 880, 890, 900, 910, 920, 930, 940] °C.
  Dataset 2 (predicted): [845, 865, 875, 875, 895, 905, 800, 925, 935, 945] °C.

  Step 1: Calculate means:
  Mean of Dataset 1 ($\bar{x}$): $\frac{850+860+\cdots+940}{10} = 895$
  Mean of Dataset 2 ($\bar{y}$): $\frac{845+865+\cdots+945}{10} = 888$

  Step 2: Calculate covariance ($\mathrm{Cov}(x,y)$):
  $\mathrm{Cov}(x,y) = 68.0$

  Step 3: Calculate standard deviations (σ_x and σ_y):
  $\sigma_x = 25.0, \sigma_y = 40.0$

  Step 4: Calculate PCC:
  $\mathrm{PCC} = \frac{\mathrm{Cov}(x,y)}{\sigma_x \cdot \sigma_y} = \frac{68.0}{25.0 \cdot 40.0} = 0.68$

  Step 5: Interpret Results:
  A PCC of 0.68 suggests a moderate positive correlation overall, with spurious deviations (e.g., 800 °C predicted vs. 910 °C measured) potentially indicating calibration, a transient event or issues during the acquisiton of data.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500] microstrain.
  Dataset 2 (predicted): [1050, 1480, 2020, 2450, 2980, 3550, 3000, 4520, 5050, 5480] microstrain.

  Step 1: Calculate means:
  Mean of Dataset 1 ($\bar{x}$): $\frac{1000+1500+\cdots+5500}{10} = 3250$
  Mean of Dataset 2 ($\bar{y}$): $\frac{1050+1480+\cdots+5480}{10} = 3248$

  Step 2: Calculate covariance ($\mathrm{Cov}(x,y)$):
  $\mathrm{Cov}(x,y) = 103500.0$

  Step 3: Calculate standard deviations (σ_x and σ_y):
  $\sigma_x = 1500.0, \sigma_y = 1425.0$

  Step 4: Calculate PCC:
  $\mathrm{PCC} = \frac{\mathrm{Cov}(x,y)}{\sigma_x \cdot \sigma_y} = \frac{103500.0}{1500.0 \cdot 1425.0} = 0.97$

  Step 5: Interpret Results:
  A PCC of 0.97 confirms a strong positive correlation, though deviations (e.g., 3000 microstrain predicted vs. 4000 microstrain measured) highlight areas where predictions may need refinement or some digital problem that might have occurred.

8. Coefficient of Determination (R^2)

Formula: $R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{residual}}}{\mathrm{SS}_{\mathrm{total}}}$

Meaning: R^2, the Coefficient of Determination, measures how well a regression model explains the variability of the dependent variable. Unlike PCC, which quantifies linear relationships, R^2 assesses the goodness-of-fit for a predictive model. Values range from 0 to 1, with higher values indicating better fit.

Key Difference Compared to PCC:

- PCC measures the strength and direction of the linear correlation between two datasets.
- R^2 evaluates how well a regression model explains the variability of one dataset based on another.
- PCC can be used without a regression model, while R^2 is directly tied to the performance of a regression model.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5] MPa.
  Regression Model Predictions: [2.9, 3.4, 4.2, 4.6, 5.2, 6.0, 6.1, 6.5, 6.9, 9.0] MPa.

  Step 1: Calculate $\mathrm{SS}_{\mathrm{residual}}$*:*
  $\mathrm{SS}_{\mathrm{residual}} = \sum(x_i - \hat{y}_i)^2 = (3.0-2.9)^2 + (3.5-3.4)^2 + \cdots + (7.5-9.0)^2$
  $= 0.01 + 0.01 + 0.04 + 0.01 + 0.04 + 0.25 + 0.01 + 0.0 + 0.01 + 2.25 = 2.63$

  Step 2: Calculate $\mathrm{SS}_{\mathrm{total}}$*:*
  $\mathrm{SS}_{\mathrm{total}} = \sum(x_i - \bar{x})^2 = 21.5$ (as calculated earlier).

  Step 3: Calculate R^2*:*
  $R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{residual}}}{\mathrm{SS}_{\mathrm{total}}} = 1 - \frac{2.63}{21.5} = 0.877$

  Step 4: Interpret Results:
  An R^2 of 0.877 indicates that the regression model explains 87.7% of the variability in the measured discharge pressure data. However, the deviation at the last data point (9.0 MPa predicted vs. 7.5 MPa measured) highlights a potential outlier or issue with the model in high-pressure ranges.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [850, 860, 870, 880, 890, 900, 910, 920, 930, 940] °C.
  Regression Model Predictions: [840, 855, 875, 885, 885, 905, 915, 930, 945, 960] °C.

  Step 1: Calculate $\mathrm{SS}_{\mathrm{residual}}$*:*
  $\mathrm{SS}_{\mathrm{residual}} = (850-840)^2 + (860-855)^2 + \cdots + (940-960)^2$
  $= 100 + 25 + 25 + 25 + 25 + 25 + 25 + 100 + 225 + 400 = 975$

  Step 2: Calculate $\mathrm{SS}_{\mathrm{total}}$*:*
  $\mathrm{SS}_{\mathrm{total}} = 8250$ (as calculated earlier).

  Step 3: Calculate R^2*:*
  $R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{residual}}}{\mathrm{SS}_{\mathrm{total}}} = 1 - \frac{975}{8250} = 0.882$

  Step 4: Interpret Results:
  An R^2 of 0.882 indicates that the regression model explains 88.2% of the variability in the turbine temperature data. Discrepancies at low and high temperatures (e.g., 840 °C predicted vs. 850 °C measured) suggest areas for improvement.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500] microstrain.
  Regression Model Predictions: [1100, 1400, 2100, 2400, 3200, 3600, 3900, 4600, 5100, 5400] microstrain.

  Step 1: Calculate $\mathrm{SS}_{\mathrm{residual}}$*:*
  $\mathrm{SS}_{\mathrm{residual}} = (1000-1100)^2 + (1500-1400)^2 + \cdots + (5500-5400)^2$
  $= 10000 + 10000 + 10000 + 10000 + 40000 + 10000 + 10000 + 10000 + 10000 + 10000 = 140000$

  Step 2: Calculate $\mathrm{SS}_{\mathrm{total}}$*:*
  $\mathrm{SS}_{\mathrm{total}} = 5062500$ (as calculated earlier).

  Step 3: Calculate R^2*:*
  $R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{residual}}}{\mathrm{SS}_{\mathrm{total}}} = 1 - \frac{140000}{5062500} = 0.972$

  Step 4: Interpret Results:
  An R^2 of 0.972 confirms a strong fit overall but highlights potential under- or overpredictions (e.g., 1100 microstrain predicted vs. 1000 microstrain measured) in some ranges.

9. Maximum Correlation Coefficient

Formula: $\text{Max Correlation} = \max\left(\frac{1}{n}\sum_{i=1}^{n-k}\frac{(x_i-\bar{x})(y_{i+k}-\bar{y})}{\sigma_x \cdot \sigma_y}\right)$

Lag at Maximum Correlation: $\text{Lag} = \text{Index of Max Correlation} - (n-1)$

Meaning: The Maximum Correlation Coefficient identifies the strongest linear relationship between two datasets across all possible time lags k. The lag at maximum correlation indicates the time offset where the strongest alignment occurs, which is particularly useful in systems with delays or shifts in signals.

Practical Use Cases:

- **Compressor Discharge Pressure:**
  Dataset 1 (measured): [3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5] MPa.
  Dataset 2 (predicted): [3.1, 3.6, 4.1, 4.4, 5.1, 5.6, 6.1, 6.6, 7.1, 7.6] MPa.

  Step 1: Normalize the datasets:
  $x_{\mathrm{norm}} = \frac{x-\bar{x}}{\sigma_x}, \quad y_{\mathrm{norm}} = \frac{y-\bar{y}}{\sigma_y}$

  Step 2: Compute cross-correlation:
  $\text{Cross Corr}(k) = \frac{1}{n}\sum_{i=1}^{n-k} x_{\mathrm{norm}}(i) \cdot y_{\mathrm{norm}}(i+k)$

  Step 3: Identify maximum correlation and lag:
  Maximum correlation: $\text{Max Correlation} = 0.98$
  Lag at max correlation: $\text{Lag} = 0$ (datasets are perfectly aligned).

  Step 4: Interpret Results:
  The maximum correlation of 0.98 occurs with no time lag, confirming excellent alignment between measured and predicted datasets under steady-state conditions.

- **Gas Turbine Inlet Temperature:**
  Dataset 1 (measured): [850, 860, 870, 880, 890, 900, 910, 920, 930, 940] °C.
  Dataset 2 (predicted): [845, 855, 865, 875, 885, 895, 905, 915, 925, 935] °C.

  Step 1: Normalize the datasets:
  Normalize as above.

  Step 2: Compute cross-correlation:
  Cross-correlation values:
  $\mathrm{Corr}_0 = 0.95, \mathrm{Corr}_1 = 0.80, \mathrm{Corr}_{-1} = 0.85$

  Step 3: Identify maximum correlation and lag:
  Maximum correlation: $\text{Max Correlation} = 0.95$
  Lag at max correlation: $\text{Lag} = 0$

  Step 4: Interpret Results:
  The maximum correlation of 0.95 with no time lag suggests good alignment overall, but lower correlations for $k = 1$ and $k = -1$ indicate reduced alignment under transient conditions.

- **Blade Strain Testing:**
  Dataset 1 (measured): [1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500] microstrain.
  Dataset 2 (predicted): [1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000] microstrain (shifted by 1 index).

  Step 1: Normalize the datasets:
  $x_{\mathrm{norm}} = \frac{x-\bar{x}}{\sigma_x}, \quad y_{\mathrm{norm}} = \frac{y-\bar{y}}{\sigma_y}$
  For measured (x): $\bar{x} = 3250, \quad \sigma_x = 1581.1$
  For predicted (y): $\bar{y} = 3750, \quad \sigma_y = 1581.1$

Step 2: Compute cross-correlation:
At $k = 0$ (no lag):

$$\text{Corr}_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \cdot \sigma_y}$$

$$= \frac{1}{10}\left(\frac{(1000 - 3250)(1500 - 3750)}{1581.1^2} + \cdots + \frac{(5500 - 3250)(6000 - 3750)}{1581.1^2}\right)$$

$$\text{Corr}_0 = 0.85$$

At $k = 1$ (predicted dataset shifted forward by 1 index):

$$\text{Corr}_1 = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{(x_i - \bar{x})(y_{i+1} - \bar{y})}{\sigma_x \cdot \sigma_y}$$

$$\text{Corr}_1 = 0.92$$

At $k = -1$ (predicted dataset shifted backward by 1 index):

$$\text{Corr}_{-1} = 0.75$$

Step 3: Identify maximum correlation and lag:
Maximum correlation: $\text{Max Correlation} = 0.92$
Lag at max correlation: $\text{Lag} = 1$

Step 4: Interpret Results:
The maximum correlation of 0.92 occurs at a lag of 1, indicating that the predicted dataset aligns best when delayed by 1 time step. This suggests a possible delay in the strain measurement model, highlighting areas where predictions may need adjustment to account for dynamic loading scenarios.