MARIN

MEMO

To	:	**Hanna Dąbrowska | Water Revolution Foundation**
From	:	**Kirsten Odendaal**
CC	:	**Enrico Della Valentina**
Date	:	**2024-09-04**
Project No.	:	
Subject	:	**YETI Reference Lines**

The General reference line development process is as follows:

1. Considers only the Eco-points metric:
    a. Based on the General Assembly (GA) meeting where the Eco-score metric demonstrated unfavourable (noisy) outcomes.
2. Eco-points versus Length and Eco-points versus GT are considered:
3. Non-linear power relations ($y = A \cdot x^B$) are only considered:
    a. Many alternatives exist; however, these functions are flexible and easy to explain.
4. Synthetic data and Real data fitting lines are compared globally:
    a. The combined fitting curves (Synthetic and real) don't deviate much from the real data. Thus, the combined datasets are used to establish the Reference line to allow for additional data.
5. prediction intervals for 1 standard deviation are evaluated:
    a. A linearization process is considered to determine the 1 standard deviation (1-STD) boundary, where prediction bounds fit in the linear domain and translate back to the power domain. Approximately 68% of the dataset population falls within these bounds.
6. A segmented case based on GT is considered:
    a. An additional case is explored, and individual reference lines are created when considering the regulatory conventions for GT.
        i. < 500GT
        ii. < 3000GT
        iii. < 10000GT (insufficient data to create a dedicated reference)

The general regression fitting notation and formulations are as follows,

Power fitting function:

$$y = A \cdot x^B$$

Power function linearization:

$$\ln(y) = \ln(A) + B * \ln(x) \quad \rightarrow \quad \underline{\boldsymbol{Y}} = \underline{\boldsymbol{A}} + B * \underline{\boldsymbol{X}}$$

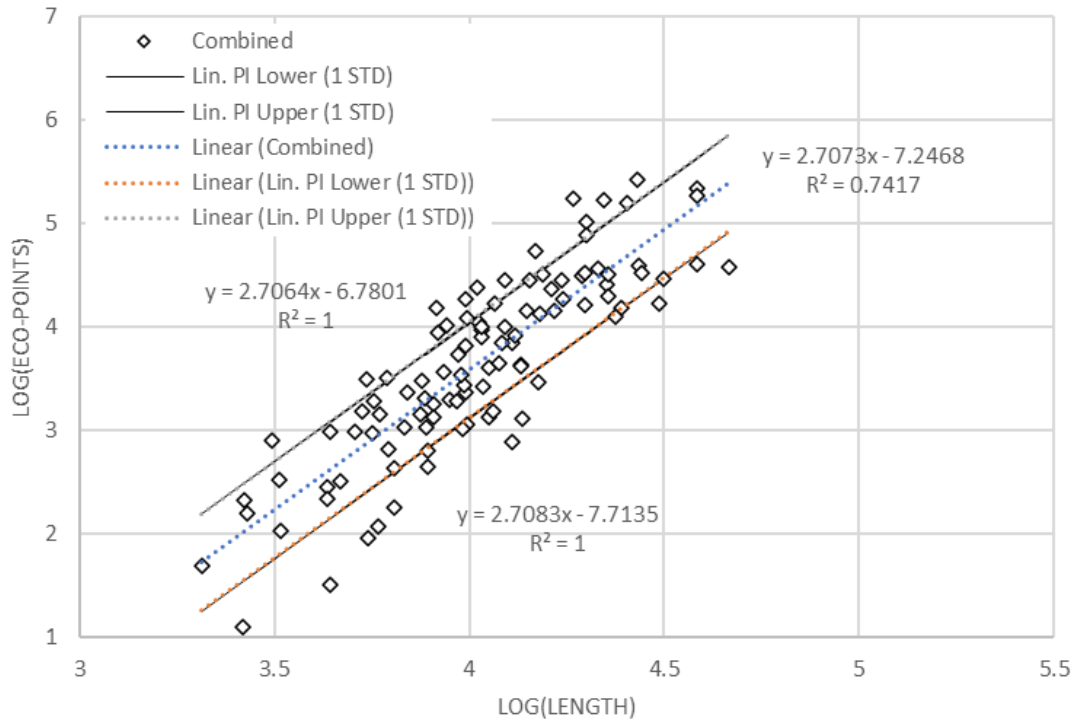
68% Prediction bounds - 1 standard deviation (1-STD):

$$\mu \pm t_{\alpha/2} * SE * \sqrt{1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$$

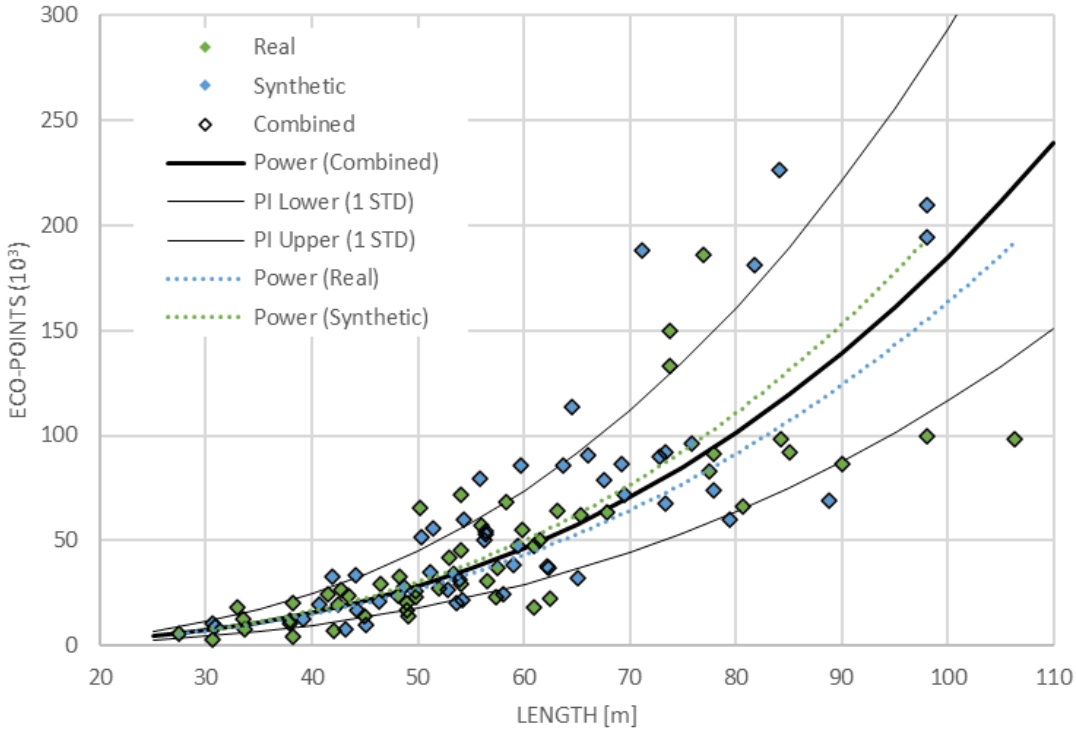
$$SE = \sqrt{\frac{\Sigma(y_i - \hat{y})^2}{n - 2}}$$

1. Eco-points versus Length:

Linearization:



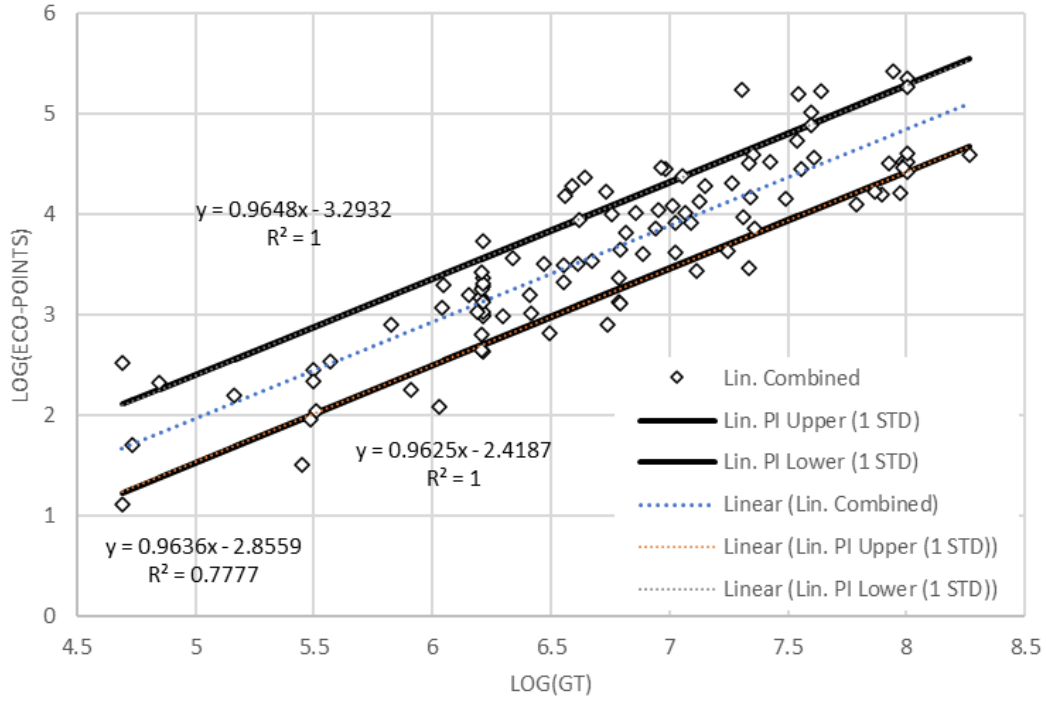
Reference Line:



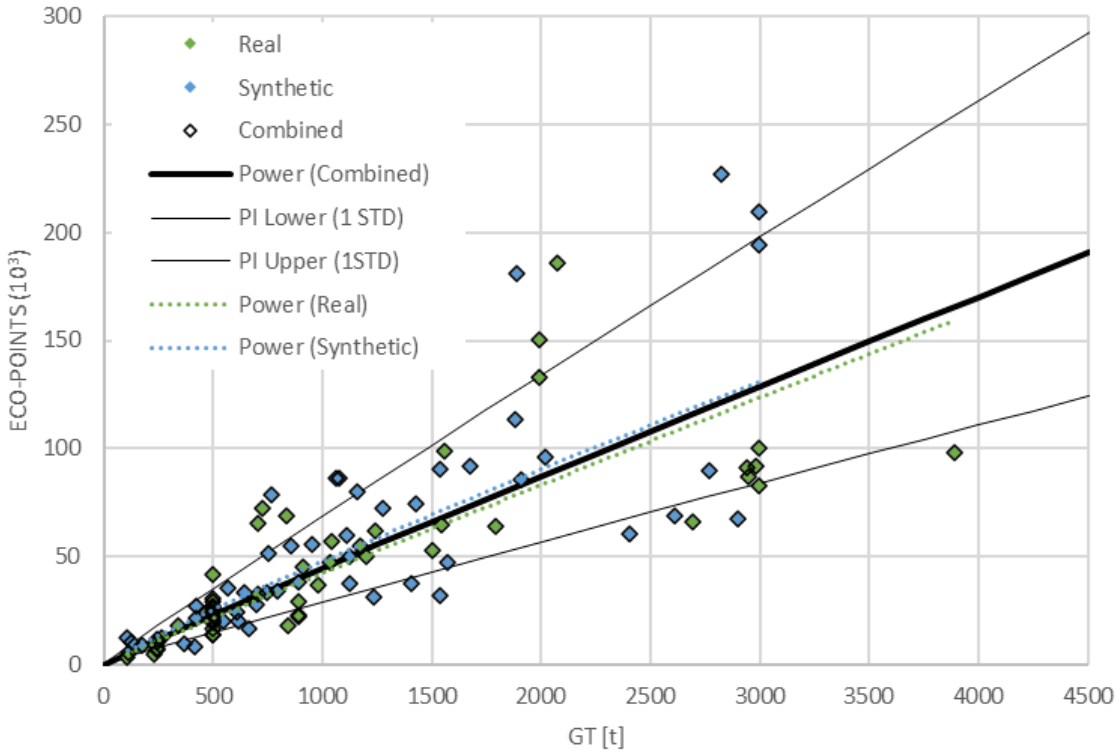
Reference line ($y = A * x^B$):	1 STD lower bound :	1 STD upper bound:
A = 0.00071	A = 0.00114	A = 0.00045
B = 2.70733	B = 2.70641	B = 2.70825

2. Eco-points versus GT:

Linearization:



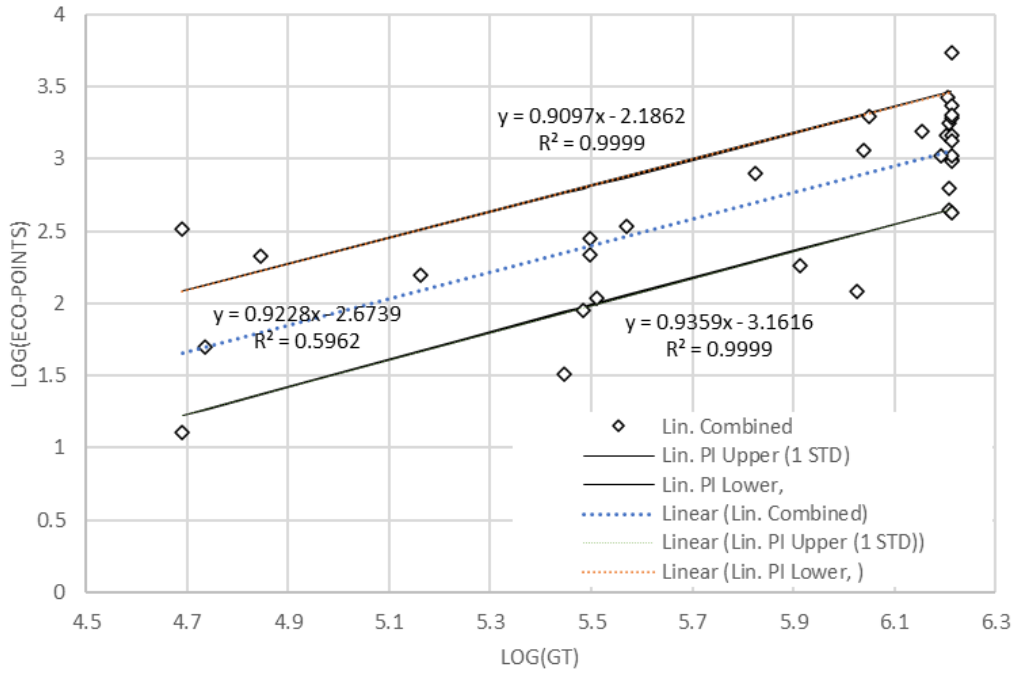
Reference Line:



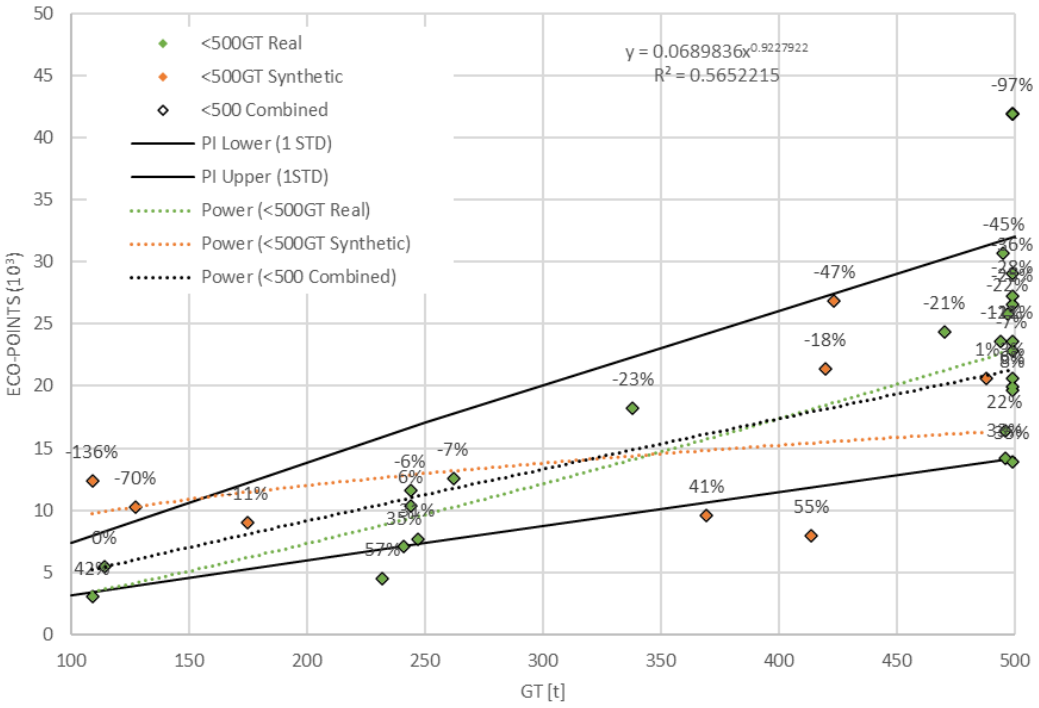
Reference line ($y = A * x^B$):	1 STD lower bound:	1 STD upper bound:
A = 0.0575013	A = 0.03713	A = 0.08904
B = 0.9636271	B = 0.964750	B = 0.962504

3. Segmented Eco-points versus GT (<500GT)

Linearization:



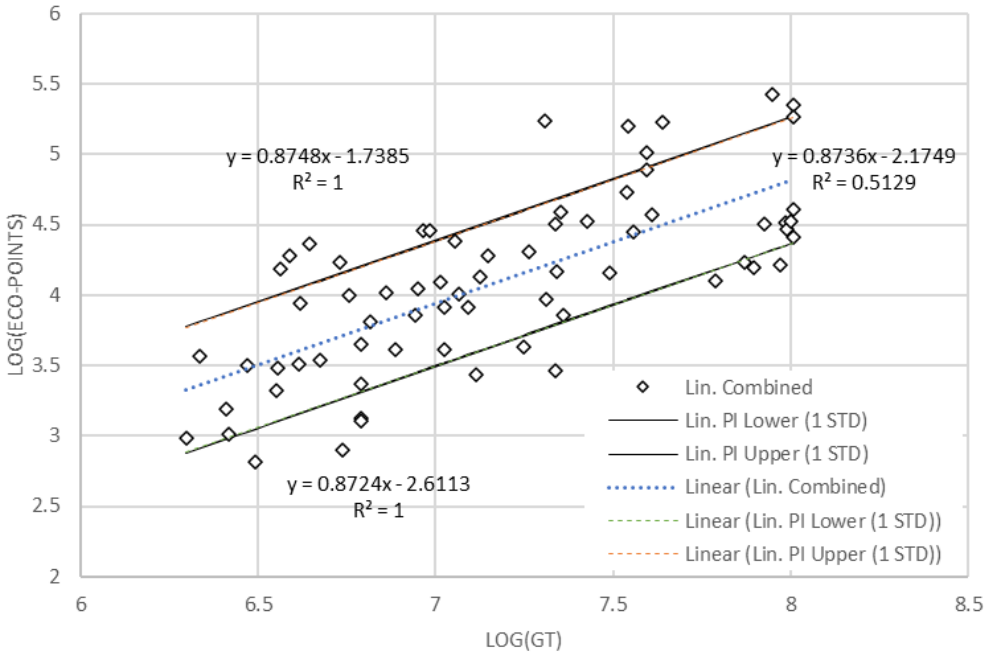
Reference Line:



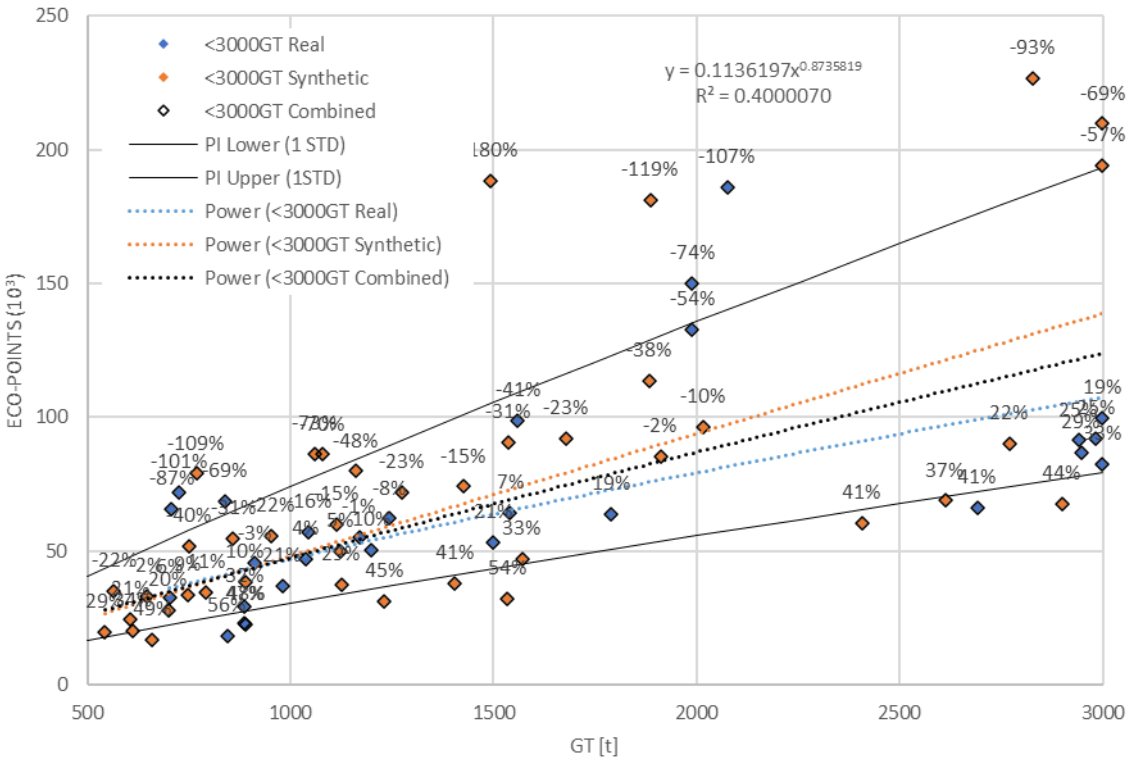
Reference line ($y = A * x^B$):	1 STD lower bound:	1 STD upper bound:
A = 0.06898	A = 0.04236	A = 0.11234
B = 0.92279	B = 0.935876	B = 0.909708

4. Segmented Eco-points versus GT (500GT< x <3000GT)

Linearization:

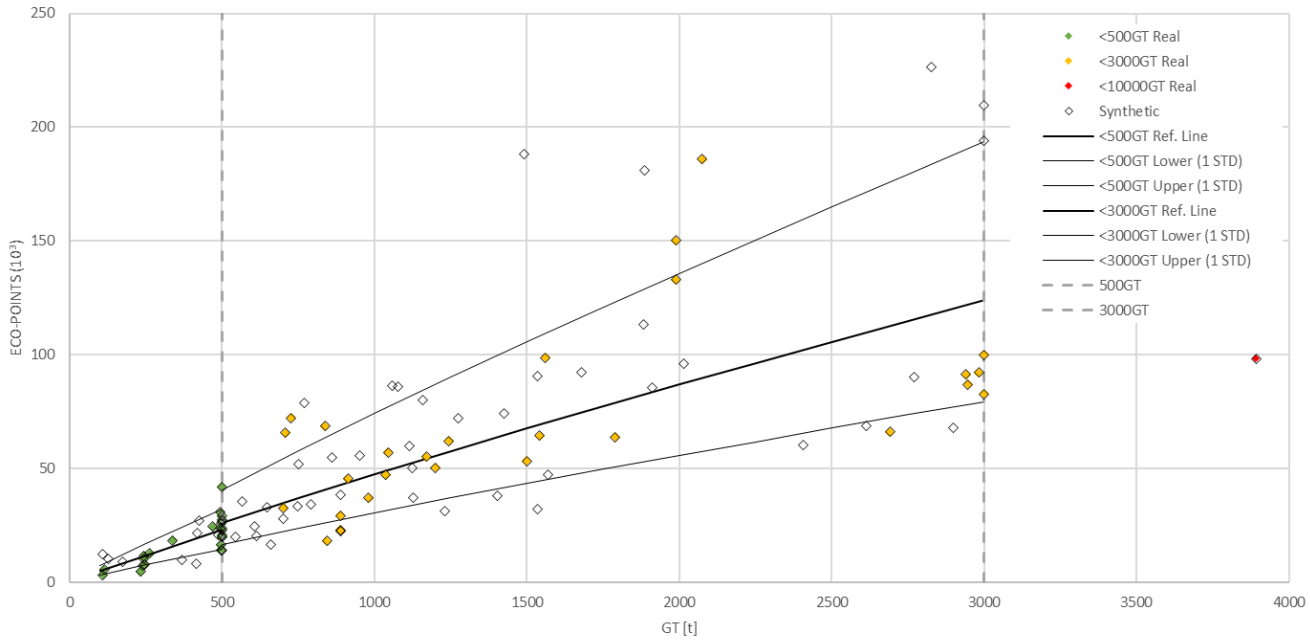


Reference Line:



Reference line ($y = A * x^B$):	1 STD lower bound:	1 STD upper bound:
A = 0.11362	A = 0.07344	A = 0.17578
B = 0.87358	B = 0.872358	B = 0.874782

Final Proposal:



Recommendations:

1. The final proposal (segmented GT reference lines) should be considered as these curves demonstrate good fitting and reflect the overall GT regulatory constraints for various yacht classes. Thus, this is the most complete representation.

2. Unfortunately, only one data point is available within the dataset that falls in the <10000GT. Thus, a reference line in this region cannot be accurately developed. Nevertheless, this proposed methodology can be applied when more data is available.

3. Due to the exponential nature of the power function, the *Length* parameter is likely a poor candidate for the reference line. As the vessel's length increases, an exponential increase in the mean is witnessed. This relation fits well with the total data available; however, it can be seen that the real data points in the upper length regions actually lay quite low. Thus, without more information in the upper length regions, there exists a large risk that larger vessels are given larger leeway for easy improvement.

4. Anything below the mean reference line predicts an improvement on the mean sample size. Anything below the second line (lower curve) demonstrates an improvement over 1 standard deviation of the data distribution (~68 percent of the data points are within the boundaries).

5. While the curves for the Eco-points vs. GT reference lines appear to be linear, They are NOT (non-linear power relations are considered). Nevertheless, they are very close to being linear and as such, linear lines could be considered and compared.