The Effect of Different Machine Learning Surrogate Models for Bayesian Optimization on the Thrust-to-Power Ratio of Physical Electrohydrodynamic Ion Thrusters

Avitej Akula, Elliott K. Choi, Gurpreet Juneja
Adlai E. Stevenson High School, Lincolnshire, Illinois, 60069, USA

ABSTRACT

Humanity continues to be constrained by terrestrial resource requirements, necessitating decarbonized propulsion for space and atmospheric applications. Ion thrusters are known to achieve 10 times the efficiency of traditional rocket engines; moreover, atmospheric electrohydrodynamic (EHD) ion thrusters offer a thrust-to-power ratio 55 times that of turbofan aircraft engines. Despite this potential, EHD thrusters are insufficient for extended atmospheric flight, especially as current methods for traditional simulation-based thruster optimization remain extremely time intensive. This study investigates Bayesian Optimization as an engineering tool to accelerate ion thruster optimization, specifically on surrogate model selection. An EHD thruster was constructed and continuously calibrated to generate an empirical dataset of 7,222 datapoints, then preprocessed through methods of standardization and outlier removal to reduce selection bias. An XGBRegressor (R-Squared = 97%, MAE = 0.04 N/kW) was then trained to serve as the objective function, and the surrogate models Gaussian Process (GP), Random Forest (RF), and Extra Trees (ET) were tested by utilizing a weighted sum that combines optimization time and thrust-to-power ratio into a single score, with both metrics given equal weights. ETR had the most balance between thrust-to-power ratio and optimization time, with 0.28 N/kW and 8.98 seconds respectively. This resulted in a weighted sum of 0.777 ($p < 0.001$, $F(2, 297) = 1737$), significantly outperforming GP and RF on weighted sum. From this study, ET's proficiency as a surrogate model reduces the reliance on exhaustive methods of manual parameter sweeping and facilitates testing of broader variables with fewer iterations, significantly accelerating development of high-performing ion thrusters.

Keywords: electrohydrodynamic ion thruster, Bayesian optimization, surrogate model, Extra Trees Regression, thrust-to-power ratio, Gaussian Process Regression

INTRODUCTION

Altitudes extending above 20km above Earth's surface are critical for exploring, meteorological research, delivery of payloads, etc. However, most chemical propulsion systems at this altitude often are inefficient, suffering in terms of efficiency due to the low atmospheric pressure (Khomich et al., 2020). Electric ion thrusters are a highly efficient form of propulsion, able to be used within deep space travel without chemical combustion (Anil, 2017). Electrohydrodynamic ion thrusters, a type of electric propulsion system (EPS), uses the principle of Electrohydrodynamic ionization to produce thrust (Iranshahi et al., 2024; Goebel & Katz, 2008). EHD ion thrusters are able to offer High Altitude Long Endurance (HALE) aircraft a sustainable and renewable form of propulsion (Khomich et al., 2020). Ion thrusters are traditionally optimized through parameter sweeps (Goebel & Katz, 2008), high-fidelity Computational Fluid Dynamics (CFD) simulations, or machine learning-based search algorithms. Yet, these methods require substantial time and resources to execute a single evaluation (Alizadeh et al., 2020). Bayesian Optimization offers a promising alternative by reducing the number of these evaluations whilst delivering equally optimized results. However, the performance of Bayesian Optimization is heavily reliant on the type of surrogate model, and the effect of different surrogate models on ion thruster optimization has not been adequately identified. The purpose of this study is to investigate Bayesian Optimization as an engineering tool for the acceleration of ion thruster thrust-to-power optimization, specifically focusing on surrogate model selection.

MATERIALS & METHODS

Phase 1, Ion Thruster Creation. A needle-to-ring EHD ion thruster was constructed using stainless steel needle anodes and a copper ring cathode (1 cm inner radius, 2 cm outer radius) similar to Zhang et al (2015). Electrode movement was controlled via a NEMA 17 stepper motor paired with a DRV8825 motor driver. Thrust was measured using a 100g load cell amplified by an HX711. High voltage was supplied through a 50 kV step-up transformer module driven by an adjustable 0–30V bench power supply, monitored by an INA226 module. All components were interfaced with an ELEGOO Uno microcontroller.

To ensure voltage accuracy, reference voltages were established using sphere-gap measurement following the Bureau of Indian Standards IS 1876 (2005), equivalent to IEC 60052. A correction factor of 0.99334908 was applied to standard atmospheric conditions to determine spark gap distances at target kilovoltages. Data was collected across voltage ranges of 2.78–9.53 kV, inter-electrode distances of 0–30 mm, and needle counts of 1–9 pins. Calibration datapoints were cross referenced to Zhang et al (2015) to maintain legitimacy. The final empirical dataset of 7,222 datapoints was saved in CSV format.

Phase 2, Objective Function Training. After this, the empirical dataset was preprocessed prior to model training. Rows containing negative or outlier thrust-to-power ratios were removed as outliers, and duplicate ratio values were removed to reduce selection bias. The pin count column was one-hot encoded as a categorical variable, and polynomial features were added to capture nonlinear relationships. All

features were scaled using scikit-learn's StandardScaler.

The preprocessed dataset was split into training, validation, and test sets at a 70:20:10 ratio. An XGBRegressor was trained with 1,100 estimators, a learning rate of 0.08, and a maximum tree depth of 10, using the validation set to monitor for overfitting. The final model achieved an R^2 of 0.967 and an MAE of 4.16×10^{-5} N/W on the test set, with validation MAE consistent with test MAE, confirming generalizability. The trained model was saved and used as the objective function for all subsequent Bayesian Optimization trials.

Phase 3, Surrogate Model Testing. Three surrogate models were evaluated within a Bayesian Optimization framework: Gaussian Process Regression (GPR), Random Forest Regression (RFR), and Extra Trees Regression (ETR). All models were tested under identical conditions using the scikit-optimize library, with a fixed search space, 50 optimization iterations, Upper Confidence Bound (UCB) as the acquisition function, and Latin Hypercube Sampling for initial points. Random state was set to the current trial index to ensure reproducibility across trials.

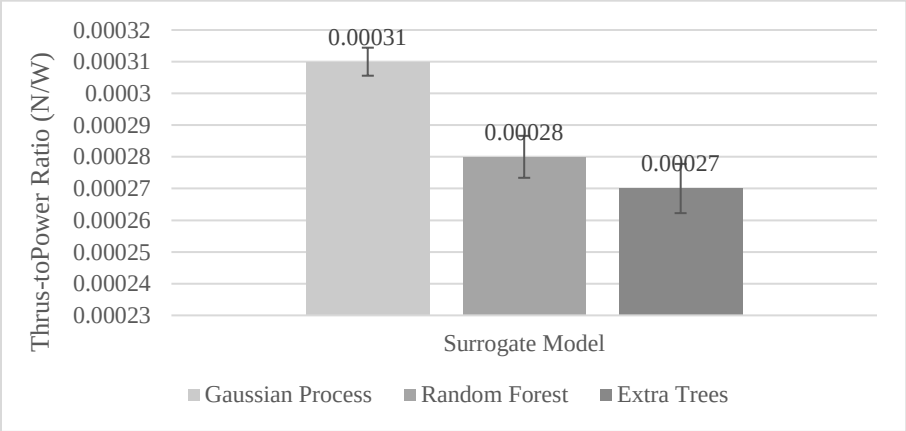

Figure 1. *The Effect of Different Surrogate Models on the Thrust-to-Power Ratio*

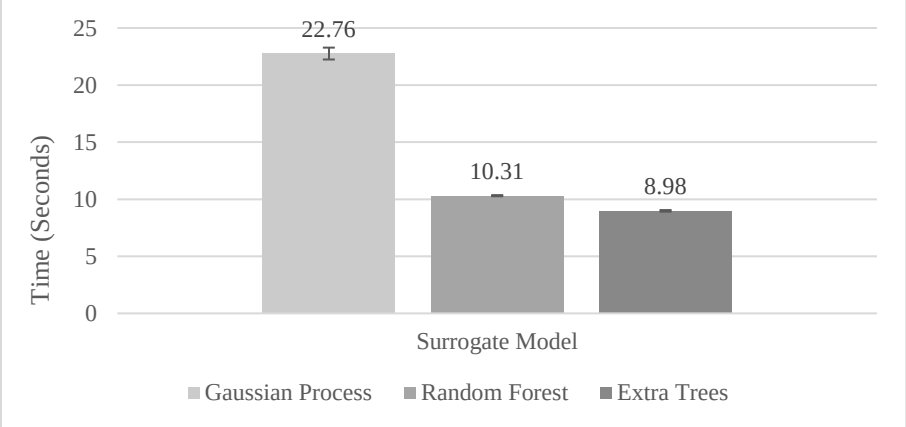

Figure 2. *The Effect of Different Surrogate Models on the Optimization Time*

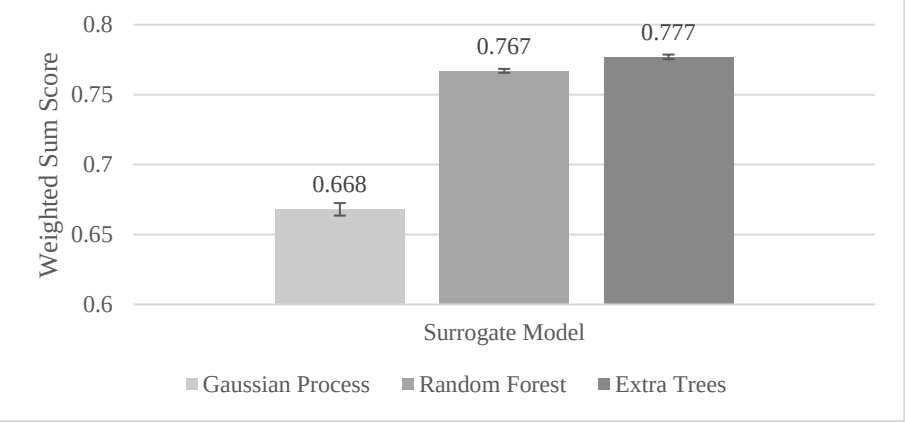

Figure 3. *The Effect of Different Surrogate Models on Weighted Sum Score*

Each surrogate model was run for 100 independent trials. For each trial, optimization time was recorded using Python's time module, and the highest thrust-to-power ratio identified was extracted. A 60-second timeout was applied to prevent indefinite runtime.

Trial outcomes were evaluated using a weighted sum method. Thrust-to-power ratio values were normalized using a logarithmic transformed min-max normalization to account for the large range of values, while optimization time was normalized using standard min-max normalization with directional flipping to reward faster convergence. Both metrics were assigned equal weights, and the weighted sum for each trial was calculated as:

$$A_{sum} = 0.5R_p' + 0.5t_i'$$

where R_p' is the normalized thrust-to-power ratio and t_i' is the normalized optimization time.

A one-way ANOVA was conducted within JASP 0.96.0 to examine the effect of different surrogate models on thrust-to-power ratio, optimization time, and weighted sum across 100 optimization trials per model. Post-hoc comparisons were performed using Tukey's HSD, with effect sizes reported as ω^2, which is bias-corrected (Kroes & Finley, 2023); η^2 is retained for transparency, given its tendency toward overestimation (Lakens, 2013). Pairwise effect sizes were interpreted using Cohen's (1988) benchmarks of small (d = 0.2), medium (d = 0.5), and large ($d \geq 0.8$). Although Levene's test indicated heterogeneity of variances across groups, the balanced design (N = 100 per group) rendered the ANOVA

F-test robust to this violation, as established by Glass et al. (1972).

RESULTS

Descriptive statistics across all three performance metrics are summarized in Table 1, one-way ANOVA results in Table 2, and post-hoc pairwise comparisons (Tukey HSD) in Table 3. For all the performance metrics, the null hypothesis (H_0: no difference in surrogate model performance, with any observed variation due to chance) was rejected at $p < 0.001$. All analyses were conducted in JASP 0.96.0.

GPR achieved the highest mean thrust-to-power ratio ($M = 3.061\times10^{-4}$ N/W, $SD = 2.240\times10^{-5}$) and the lowest variability (CV = 7.3%), followed by RFR ($M = 2.835\times10^{-4}$ N/W, CV = 11.8%) and ETR ($M = 2.787\times10^{-4}$ N/W, CV = 14.1%). Surrogate model type had a significant effect on thrust-to-power ratio ($F(2, 297) = 20.30$, $p = 5.435\times10^{-9}$, $\omega^2 = 0.114$, 95% CI [0.052, 0.182]), with GPR significantly outperforming both RFR ($p = 4.114\times10^{-6}$, $d = 0.697$) and ETR ($p = 2.112\times10^{-8}$, $d = 0.843$). No significant difference was observed between RFR and ETR ($p = 0.556$).

ETR achieved the fastest mean optimization time ($M = 8.98$ s, CV = 3.9%), followed by RFR ($M = 10.31$ s, CV = 2.4%) and GPR ($M = 22.76$ s, CV = 11.6%). Surrogate model type had a large effect on optimization time ($F(2, 297) = 2427$, $p = 9.955\times10^{-185}$, $\omega^2 = 0.942$, 95% CI [0.931, 0.950]), with all pairwise comparisons significant and practically large (ETR vs. GPR: $d = -8.932$; GPR vs. RFR: $d = 8.067$; ETR vs. RFR: $d = -0.865$).

ETR yielded the highest mean weighted sum ($M = 0.777$, $SD = 0.008$), followed by RFR ($M = 0.767$, CV = 0.9%) and GPR ($M = 0.668$, CV = 3.4%). Surrogate model type had a large effect on weighted sum ($F(2, 297) = 1737$, $p = 1.257\times10^{-164}$, $\omega^2 = 0.920$, 95% CI [0.906, 0.932]), with all pairwise comparisons statistically significant and practically meaningful (ETR vs. GPR: $d = 7.540$; GPR vs. RFR: $d = -6.848$; ETR vs. RFR: $d = 0.691$).

Table 1. *Descriptive Statistics for Surrogate Models Across Performance Metrics*

Metric	Model	N	Mean	SD	SE	CV
Thrust-to-Power	GPR	100	3.061×10^{-4}	2.240×10^{-5}	2.240×10^{-6}	0.073
	RFR	100	2.835×10^{-4}	3.353×10^{-5}	3.353×10^{-6}	0.118
	ETR	100	2.787×10^{-4}	3.928×10^{-5}	3.928×10^{-6}	0.141
Time	GPR	100	22.76	2.637	0.264	0.116
	RFR	100	10.31	0.249	0.025	0.024
	ETR	100	8.98	0.349	0.035	0.039
Weighted Sum	GPR	100	0.668	0.023	0.002	0.034
	RFR	100	0.767	0.007	7.145×10^{-4}	0.009
	ETR	100	0.777	0.008	8.258×10^{-4}	0.011

Table 2. *One-Way ANOVA Results for Surrogate Model Comparison*

Metric	F	df	p	η^2	95% CI η^2	ω^2	95% CI ω^2
Thrust-to-Power	20.3	2, 297	5.435×10^{-9}	0.12	[0.057, 0.189]	0.114	[0.052, 0.182]
Time	2427	2, 297	9.955×10^{-185}	0.942	[0.932, 0.951]	0.942	[0.931, 0.950]
Weighted Sum	1737	2, 297	1.257×10^{-164}	0.921	[0.907, 0.932]	0.92	[0.906, 0.932]

Table 3. *Tukey HSD Post-Hoc Comparisons*

Metric	Comparison	Mean Diff.	SE	t	Cohen's d	p
Thrust-to-Power	ETR − GPR	-2.741×10^{-5}	4.597×10^{-6}	−5.962	−0.843	2.112×10^{-8}
	ETR − RFR	-4.751×10^{-6}	4.597×10^{-6}	−1.034	−0.146	0.556
	GPR − RFR	2.266×10^{-5}	4.597×10^{-6}	4.929	0.697	4.114×10^{-6}
Time	ETR − GPR	−13.78	0.218	−63.16	−8.932	8.686×10^{-13}
	ETR − RFR	−1.334	0.218	−6.117	−0.865	8.994×10^{-9}
	GPR − RFR	12.44	0.218	57.04	8.067	8.686×10^{-13}
Weighted Sum	ETR − GPR	0.109	0.002	53.31	7.54	8.686×10^{-13}
	ETR − RFR	0.01	0.002	4.888	0.691	4.979×10^{-6}
	GPR − RFR	−0.099	0.002	−48.42	−6.848	8.686×10^{-13}

DISCUSSION

GPR achieved the highest mean thrust-to-power ratio of 0.0003061 N/W, with the one-way ANOVA revealing a significant effect on thrust efficiency. GPR significantly outperformed RFR's mean ratio of 0.00028 N/W. Across GPR's 100 runs, results deviated by 0.0000224 N/W, with a coefficient of variation of 7.3%, which is the lowest of the three models, demonstrating that GPR exhibits the greatest consistency for thrust efficiency calculation and optimization. This consistency is due to GPR's probabilistic structure, in which an uncertainty value is quantified analytically via the variance of the kernel function (Wang, 2023). For both RFR and ETR, the variance between each individual tree is calculated, and a higher variance results in a higher uncertainty value (Dutschmann & Baumann, 2021). GPR's method is much more mathematically exact since it takes both the data inputs and outputs into consideration, while the ensemble methods only consider the outputs. Pairwise effect sizes were large for GPR vs ETR ($d = 0.843$) and medium for GPR vs RFR ($d = 0.697$) per Cohen's benchmarks ($d \geq 0.8$) (Cohen, 1988).

Regarding the optimization time, ETR achieved the fastest time with an average of 8.98 seconds, 2.53 times faster than GPR's mean and 1.15 times faster than RFR's mean. The ANOVA ω^2 and Cohen's d indicated that the surrogate models had a substantial effect on optimization time. The 95% confidence interval for ω^2 is [0.931, 0.950], indicating a well-estimated

effect. Between each pairing of surrogate models, each was statistically significant from each other. Across all 100 samples, ETR reached a SD of 0.349 seconds and 3.9% CV, which indicated it could provide strong temporal reliability; however, RFR provided even more consistency (SD = 0.249, CV = 2.4%). GPR had a SD of 2.637 seconds and 11.6% CV, 4.83 times that of ETR, demonstrating it cannot reliably produce consistent runtimes. GPR's inferior temporal performance is due to its scalability constraints. GPR is computationally complex: as the data size increases, the time complexity increases cubically, rendering it increasingly demanding as dataset size grows (Liu et al., 2018). The time complexity of both RFR (Biau & Scornet, 2016) and ETR (Geurts et al., 2006) is logarithmic, therefore they take considerably less time. ETR's computational advantage stems from its randomized threshold selection, which bypasses the exhaustive splitting criterion optimization inherent to RFR. Notably, the Cohen's d values were exceptionally large (ETR vs GPR: $d = -8.932$, GPR vs RFR: $d = 8.067$, ETR vs RFR: $d = -0.865$), reflecting a practically significant divergence in computational cost across surrogate model frameworks.

This study hypothesized that GPR would yield the highest weighted sum. Contrary to this hypothesis, ANOVA revealed that ETR yielded the highest mean weighted sum of 0.777, suggesting the most favorable trade-off between thrust efficiency and computational cost across the three models ($F(2, 297) = 1737$, $p = 1.257\times10^{-164}$, $\omega^2 = 0.920$). The 95% confidence interval for ω^2 [0.906, 0.932] indicates a precisely estimated and robust effect. RFR demonstrated the second highest mean weighted sum of 0.767 and exhibited the lowest observed variability with a SD of 0.007 and a low CV of 0.9%, indicating RFR as the most stable surrogate in regards to the reproducibility of optimization outcomes. GPR yielded the lowest sum score with a weighted average of 0.668 and a SD of 0.023 with a CV of 3.4%, exhibiting the most variability out of the three surrogate models. Between each pairing of surrogate models, each was statistically significant from each other: ETR vs GPR ($p = 8.686\times10^{-13}$), ETR vs RFR ($p = 4.979\times10^{-6}$), and GPR vs RFR ($p = 8.686\times10^{-13}$). Finally, pairwise effect sizes were substantial for ETR vs GPR ($d = 7.540$) and GPR vs RFR ($d = -6.848$), and medium-large for ETR vs RFR ($d = 0.691$), collectively indicating meaningful practical divergence in surrogate model performance beyond statistical significance. ETR's 16.32% weighted sum advantage over GPR demonstrates that optimization efficiency, not thrust performance alone, determines surrogate suitability for Bayesian optimization of ion thrusters.

CONCLUSION

ETR proved its ability as a surrogate model by capturing the non-linear relationship between thruster operating parameters and efficiency with significantly higher balance of optimization time and optimized results compared to other algorithms like GPR and RFR, with a weighted sum of 0.777, 16.32% better than GPR. RFR's weighted sum was also higher than GPR by 14.82%. This suggests that ensemble-based methods may be better suited to the high-variance parameter spaces characteristic of plasma-driven propulsion systems than probabilistic alternatives. Collectively, Bayesian optimization provides a viable and computationally sound framework for EHD ion thruster optimization.

Based on the findings of this study, ETR is recommended as the best general-purpose surrogate model selection for Bayesian optimization of EHD ion thrusters, with the highest weighted sum and the fastest convergence time. RFR is recommended when reliability and consistency is the primary mission requirement, with the lowest CVs seen in the study, ideal for applications requiring a predictable surrogate with reproducible optimization behavior across validation cycles. GPR is recommended exclusively in scenarios where maximizing thrust-to-power ratio is the singular mission requirement and computational cost is not a limiting factor. Its mean thrust-to-power ratio is significantly higher and reflects a practically large advantage over RFR, however this comes at the cost of being 86.83% slower. In summary, surrogate selection should be governed by mission requirements: ETR excels at overall optimization, RFR excels at consistency, and GPR excels at thrust efficiency.

A fundamental limitation in Computational Fluid Dynamics (CFD) simulations is their computationally intensive nature, burdening computer systems through repeated trials of validation. This study presents Bayesian Optimization, specifically ETR as a surrogate model, as a promising alternative for accurate and timely optimization. By implementing this method, scientists can significantly accelerate the testing and validation of their own mission-specific EHD ion thruster datasets. This project studied the characteristics of EHD ion thrusters specifically, but the findings are just as applicable to non-airbreathing ion thruster EP following the same principles of ionization, namely Gridded Ion Thrusters, FEEP, or Electrospray Thrusters (Via Satellite, 2024; MIT, 2015). EHD and other ionization-based EPs are anticipated to be used in different conditions, including the Earth's upper atmosphere, Mars' sparse atmosphere, and in deep space travel (with the use of a gas propellant) where drag is negligible and input-to-output efficiency is crucial (Cermak, 2024). The 2.54 times speed advantage of ETR over GPR and its 16.3% higher weighted sum score represent meaningful engineering gains in optimization efficiency. For ion thruster design where iterative optimization may require hundreds of evaluations, the computational savings of ETR translate directly into reduced design cycle time without sacrificing solution quality. As electric propulsion systems become increasingly paramount to spacecraft design, Bayesian optimization offers a scalable and efficient pathway for accelerating thruster development cycles.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, AA and EC; methodology, AA and EC; data curation, AA and EC; validation, AA and EC; writing, AA and EC; editing AA and EC; Project Administration, Gurpreet Juneja (Adlai E. Stevenson High School); Supervision, Spandana Bodapati, Rama Akula, Mikyoung Choi, and Seunghyun Choi.

FUNDING

Awarded the 2026 Illinois Junior Academy of Science President's Grant.

ACKNOWLEDGEMENTS

We want to show our gratitude for our mentor, Ms. Juneja. She provided unwavering support, putting forward her personal time to provide aid. This project is just as much of an indication of her dedication and generosity as to our own efforts. We were able to ask for help in unscheduled, unpredictable times, and she would provide her full guidance without second thought, showing how much of an amazing mentor she is.

We also thank our parents, Spandana Bodapati, Rama Akula, Mikyoung Choi, and Seunghyun Choi deeply, as they provided financial and moral support throughout the project. Without our families, we would have given up long ago. They always reassured us that to pursue our dreams of STEM research, we should focus on understanding and learning rather than any monetary cost. We will be eternally grateful for their help.

LITERATURE CITED

Alizadeh, R., Allen, J. K., & Mistree, F. (2020). Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design, 31*(3), 275–298. https://doi.org/10.1007/s00163-020-00336-7

Anil, K. (2017). Electrostatic ION thruster for spacecrafts. *International Journal of Innovative Science and Research Technology*, November, 19–21. https://ijisrt.com/wp-content/uploads/2017/11/ELECTROSTATIC-ION-THRUSTER-FOR-SPACECRAFTS-2.pdf

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test, 25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bureau of Indian Standards. (2005). *Voltage measurement by means of standard air gaps* (IS 1876, First Revision). https://law.resource.org/pub/in/bis/S05/is.1876.2005.pdf

Cermak, A. (2024, November 3). Ion propulsion. NASA Science. https://science.nasa.gov/mission/dawn/technology/ion-propulsion/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Dutschmann, T., & Baumann, K. (2021). Evaluating high-variance leaves as uncertainty measure for random forest regression. *Molecules, 26*(21), 6514. https://doi.org/10.3390/molecules26216514

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*(3), 237–288. https://doi.org/10.3102/00346543042003237

Goebel, D. M., Katz, I., & NASA. (2008). *Fundamentals of electric propulsion: Ion and Hall thrusters.* https://descanso.jpl.nasa.gov/SciTechBook/series1/Goebel__cmprsd_opt.pdf

Iranshahi, K., Defraeye, T., Rossi, R. M., & Müller, U. C. (2024). Electrohydrodynamics and its applications: Recent advances and future perspectives. *International Journal of Heat and Mass Transfer, 232*, 125895. https://doi.org/10.1016/j.ijheatmasstransfer.2024.125895

Khomich, V., Malanichev, V., & Rebrov, I. (2020). Electrohydrodynamic thruster for near-space applications. *Acta Astronautica, 180*, 141–148. https://doi.org/10.1016/j.actaastro.2020.12.002

Kroes, A. D. A., & Finley, J. R. (2023). Demystifying omega squared: Practical guidance for effect size in common analysis of variance designs. *Psychological Methods, 30*(4), 866–887. https://doi.org/10.1037/met0000581

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Liu, H., Ong, Y., Shen, X., & Cai, J. (2018). When Gaussian process meets big data: A review of scalable GPs. *arXiv*. https://doi.org/10.48550/arxiv.1807.01065

MIT. (2015). *Session 20: Electrospray propulsion.* https://ocw.mit.edu/courses/16-522-space-propulsion-spring-2015/6fe2f497158a264687e8f962eac3db35_MIT16_522S15_Lecture20.pdf

Via Satellite, Enpulsion. (2024, July 28). FEEP: Field emission electric propulsion. *Via Satellite*, August 2024. https://interactive.satellitetoday.com/via/august-2024/feep-field-emission-electric-propulsion

Wang, J. (2023). An intuitive tutorial to Gaussian process regression. *Computing in Science & Engineering, 25*(4), 4–11. https://doi.org/10.1109/mcse.2023.3342149

Zhang, Y., Liu, L., Chen, Y., & Ouyang, J. (2015). Characteristics of ionic wind in needle-to-ring corona discharge. *Journal of Electrostatics, 74*, 15–20. https://doi.org/10.1016/j.elstat.2014.12.008