Gaussian Process For MSP of Sustainable Aviation Fuel

Rayman Angulo Gutierrez
Department Biological Systems Engineering
Washington State University
Pullman, WA 99163
rayman.angulo@wsu.edu

Gerald Hoff
Department of Computer Science
Washington State University
Pullman, WA 99163
gerald.hoff@wsu.edu

Sowmith Kunapaneni
Department of Computer Science
Washington State University
Pullman, WA 99163
sowmith.kunapaneni@wsu.edu

Abstract

This work presents a Gaussian Process (GP) surrogate model coupled with a Bayesian Optimization framework to evaluate and minimize the Minimum Selling Price (MSP) of Sustainable Aviation Fuel (SAF) produced via the Alcohol-to-Jet (ATJ) pathway. A System Dynamics model implemented in Stella Architect was employed to generate techno-economic data by varying eight key economic drivers. The resulting dataset enabled the GP surrogate to capture the nonlinear behavior of the ATJ system, demonstrating the utility of integrating dynamic process modeling with machine-learning methods for SAF cost assessment. An initial design of 50 simulations was used to train an Automatic Relevance Determination (ARD) RBF kernel, with additional points iteratively selected through a variance-based acquisition function. The surrogate reproduced MSP values with high accuracy, identified ethanol price as the dominant cost driver, and revealed a minimum MSP of 0.318 USD/L. A notable limitation was the absence of an automated interface between Stella Architect and Python, which required manual data transfer and constrained the number of optimization cycles. These findings highlight both the promise and practical challenges of applying surrogate-assisted optimization to complex techno-economic systems.

1 Introduction

Sustainable Aviation Fuel (SAF) has emerged as a critical pathway for reducing greenhouse gas emissions in the aviation sector, with techno-economic feasibility remaining a central challenge [U.S. Department of Energy]. Among the available production routes, the Alcohol-to-Jet (ATJ) pathway offers a promising option, yet its economic viability is highly sensitive to fluctuations in feedstock and energy markets. Understanding and minimizing the Minimum Selling Price (MSP) of SAF is therefore essential for guiding investment and policy decisions.

System Dynamics modeling provides a means to simulate the economic behavior of SAF production systems under varying market and policy conditions. In this study, a System Dynamics model implemented in Stella Architect [isee systems, inc., 2025] was used to represent the ATJ process and generate synthetic techno-economic data by varying ethanol price, diesel and gasoline market prices, electricity rate, natural gas cost, hydrogen price, loan interest rate, and federal tax credits. While such models capture system-level interactions, their computational demands can limit the scope of optimization.

To address this challenge, we develop a Gaussian Process (GP) surrogate model integrated with a Bayesian Optimization framework. The surrogate was initially trained on 50 simulations using an Automatic Relevance Determination (ARD) RBF kernel, and iteratively refined by selecting new sampling points through a variance-based acquisition function. This approach enables efficient exploration of high-dimensional economic spaces, progressively improving predictive accuracy while reducing reliance on exhaustive simulation. The methodology provides a scalable means of identifying key cost drivers and locating optimal operating conditions, thereby advancing the techno-economic assessment of SAF production.

2 Research Problem

Sustainable Aviation Fuel (SAF) is widely recognized as a central strategy for enabling the aviation industry to achieve net-zero carbon emissions by 2050 [U.S. Department of Energy]. Produced from non-petroleum feedstocks, SAF can be blended with conventional jet fuel and utilized in existing aircraft and infrastructure without modification. Depending on the feedstock, blending ratios range from 10% to 50%. In particular, the Alcohol-to-Jet (ATJ) pathway derived from cellulosic biomass allows blending up to 50%. SAF has already been deployed in more than 360,000 commercial flights, underscoring its practical importance and its role as a critical measure for reducing aviation-related emissions.

System Dynamics (SD) [Sterman, 2002] provides a framework for analyzing how complex systems evolve over time, emphasizing the interplay of stocks, flows, feedback loops, and delays. Stocks represent accumulations such as inventory or capital reserves, while flows capture the rates of change that increase or deplete these stocks (e.g., sales per day or production per month). Feedback loops and delays describe how the current state of the system influences decisions, which in turn reshape the system state, often with time lags that significantly alter outcomes.

This modeling approach is particularly effective when system behavior arises from accumulation, feedback, and delay features that characterize most real-world operational and economic problems. By explicitly incorporating time-dependent dynamics, SD captures how immediate responses to events can diverge substantially once delays are accounted for.

For the specific challenge of determining the Minimum Selling Price (MSP) of SAF, SD modeling is well-suited because the problem depends on variables that evolve over time, including production rates, market prices, policy incentives, and energy costs. Capturing these temporal interactions is essential for realistic economic assessment, and SD provides a robust foundation for integrating such dynamics into the analysis of SAF cost competitiveness.

3 Literature Review

Application of a Gaussian Process surrogate model to a Systems Dynamics problem is not new. There exist many examples in literature of this strategy.

Recent advances in surrogate modeling have strengthened the integration of machine learning (ML) with computationally intensive simulations. Gaussian Process (GP) regression provides a flexible nonparametric framework capable of capturing nonlinear relationships while delivering principled uncertainty quantification. Foundational and applied reviews demonstrate its effectiveness for emulating complex chemical processes and guiding simulation-based exploration Lucas et al. [2021], Rasmussen and Williams [2006], Meyer et al. [2021]. When coupled with Bayesian Optimization (BO), GP surrogates enable efficient global optimization with a limited number of evaluations, making them ideal for high-dimensional techno-economic design spaces Snoek et al. [2012], Shahriari et al. [2016]. These methodological advances establish a strong basis for data-efficient optimization in domains where simulations—such as TEA, dynamic economic modeling, or SAF cost analysis—are computationally intensive.

The integration of ML with scientific simulation has been increasingly recognized as a powerful paradigm for improving model accuracy, robustness, and efficiency. Hybrid modeling frameworks reviewed by von Rueden et al. show how simulation-based prior knowledge can be combined with data-driven learning to overcome limitations inherent in either approach alone von Rueden et al. [2020]. Similarly, Badakhshan et al. demonstrate that simulation can effectively generate structured datasets for ML tasks such as prediction, optimization, and scenario planning in complex engineered systems Badakhshan et al. [2024]. Complementing these conceptual frameworks, methodological advances in GP-based dynamical modeling—such as Gaussian Process Dynamical Models (GPDMs) and Variational Gaussian Process Dynamical Systems (VGPDS)—demonstrate how GP priors can capture nonlinear temporal behavior in high-dimensional systems Wang et al. [2005], Damianou et al. [2011]. These works reinforce the suitability of GP-based surrogates for representing dynamic techno-economic systems such as biofuel markets.

Despite these advances, applications of GP–BO and ML–simulation hybrid approaches in Sustainable Aviation Fuel (SAF) techno-economic assessment remain limited. Existing ATJ-focused studies rely primarily on process models and static techno-economic analysis to estimate the Minimum Selling Price (MSP), as seen in the work of Yao et al. Yao et al. [2017], Geleynse et al. Geleynse et al. [2018], and Brandt and Wolcott Brandt and Wolcott [2022]. These studies offer valuable baseline cost assessments but do not apply surrogate modeling or Bayesian optimization to systematically explore nonlinear interactions among economic drivers. As highlighted by the broader ML–simulation literature, surrogate-based optimization can dramatically improve computational efficiency and insight generation, yet no published SAF work has integrated simulation-generated TEA data with GP surrogate modeling and BO to identify MSP-minimizing conditions. This gap motivates the present study.

4 Methodology

The workflow consists of three main stages:

1. Data generation using Stella Architect,
2. Gaussian Process (GP) surrogate training, and
3. Iterative Bayesian Optimization.

This methodology is shown graphically as follows. We begin with the data generation using Stella, then implement the Gaussian Surrogate model, and iteratively apply the acquisition function.

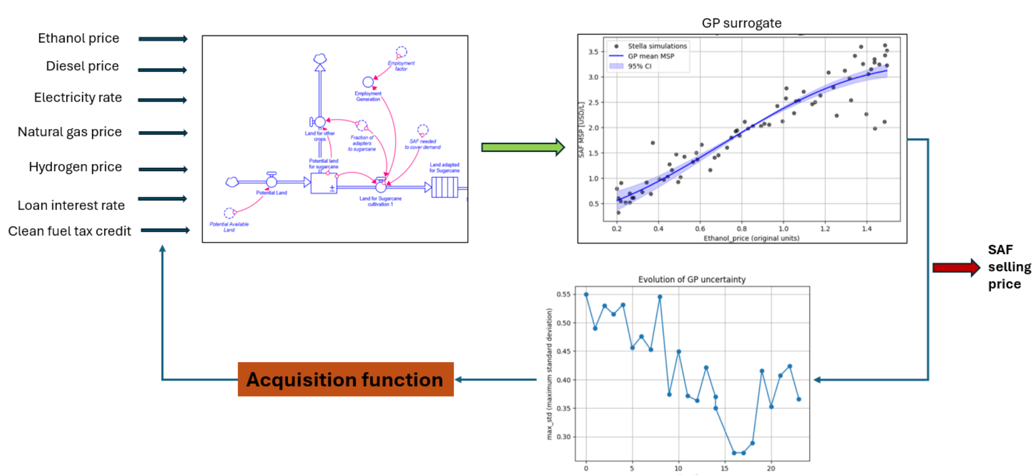

Figure 1: Methodology Abstract

This methodology is executed through google Colab using a simple CPU runtime. The process is computationally inexpensive and should be simple to replicate using the given data and code.

4.1 Data Generation Using Stella Architecture

The ATJ techno-economic system was simulated using a System Dynamics model developed in Stella Architect by Paulina Echeverria Paredes, a PhD student in Biological Systems Engineering at Washington State University. The SD model itself is not yet published.Echeverria Paredes [2025] An initial experimental design of 50 simulations was constructed by sampling eight economic input variables: ethanol price, diesel price, gasoline price, electricity rate, natural gas price, hydrogen price, loan interest rate, and the federal clean-fuels tax credit. Model outputs were exported and processed with a custom R script that performed unit conversions, cleaned variable names, removed duplicates, and formatted the dataset for machine-learning readiness.

4.2 Gaussian Process Surrogate Model Training

A Gaussian Process Regression surrogate with an Automatic Relevance Determination (ARD) RBF kernel was trained using standardized inputs and outputs. Kernel hyper-parameters were optimized by maximizing the log-marginal likelihood. Model performance was assessed using the coefficient of determination (R^2) and root-mean-square error (RMSE), and a full-data fit was used to validate the GP's ability to reproduce the System Dynamics MSP estimates.

4.3 Bayesian Optimization Process

A variance-based acquisition function was employed to guide the exploration of the input space. At each iteration, $5,000$ candidate points were uniformly sampled within predefined economic bounds, MSP uncertainty was predicted using GP, and the candidate with the highest predictive standard deviation (excluding duplicates) was selected for re-simulation in Stella. After each new simulation, the GP surrogate was retrained, generating learning curves for model error and epistemic uncertainty. This iterative process progressively expanded the sampled domain and reduced global predictive uncertainty.

5 Results

5.1 GP training

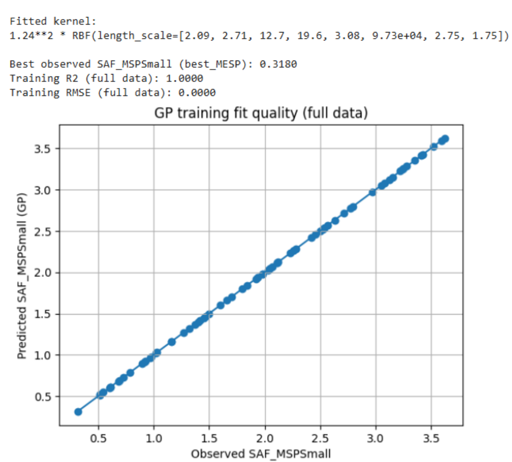

Figure 2: GP training fit (observed vs. predicted)

The scatter plot shows an almost perfect one-to-one correspondence between the observed and GP-predicted values of the MSP reflected in the training metrics $R2 = 1.0000$ and $RMSE \approx 0$. This indicates that the GP surrogate fully interpolates the available simulation data, as expected for Gaussian Processes trained with very small noise and smooth kernels. The fitted kernel hyper-parameters suggest anisotropic sensitivity across input dimensions, but the model captures the underlying mapping without residual error within the sampled domain.

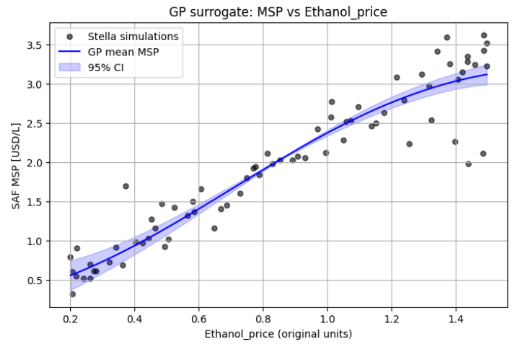

Figure 3: GP Surrogate Response Curve for SAF MSP as a Function of Ethanol Price

The figure shows the one-dimensional GP response of SAF minimum selling price (MSP) with respect to Ethanol_price. The GP mean closely follows the trend of the simulation data, capturing the nonlinear relationship between feedstock price and SAF production cost. The 95% confidence interval widens at the extremes of the input domain, reflecting higher epistemic uncertainty where fewer samples are available. Overall, the surrogate model provides a smooth and well-calibrated approximation of the underlying process behavior.

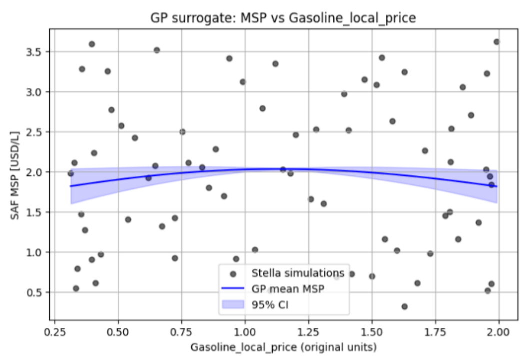

Figure 4: GP Surrogate Response: SAF MSP vs. Gasoline Local Price

The GP surrogate shows a weak, nearly flat dependence of SAF MSP on `Gasoline_local_price`, indicating that gasoline price has minimal influence on the modeled MSP across the explored range. The confidence interval widens slightly at the domain boundaries, reflecting higher uncertainty where fewer simulation points are available. The scatter of Stella simulations confirms the low sensitivity, with substantial variability but no strong monotonic trend. The same was observed for the other inputs, just ethanol has a important effect on the price of the SAF.

This plot shows how the Gaussian Process (GP) surrogate model's root-mean-square error (RMSE) evolves as the number of training points increases. The RMSE increases steadily from approximately 2.0×10^{-6} at 10 training points to a peak of about 5.5×10^{-6} around 65 points, with a slight decrease near 75 points. This upward trend indicates that, as additional data points are included, particularly points selected by the acquisition function in regions of previously high uncertainty, the GP must fit a more complex response surface, leading to higher training error. This behavior is expected: adding exploratory samples reduces overfitting to the initial dataset and forces the model to generalize across a broader domain, which naturally increases the training residuals.

The R^2 curve exhibits the complementary trend: as the number of training samples grows, the training R^2 decreases from values near 0.96 down to approximately 0.68 at 65 points, with a moderate recovery near 75 points. The monotonic decline in R^2 reflects the same phenomenon observed in the RMSE curve: additional exploratory samples introduce new variability and reduce the GP's ability to perfectly interpolate the initial dataset. This reduction in training R^2 is typical for GP models undergoing active exploration, where the acquisition function (max-variance sampling) intentionally prioritizes regions of maximal uncertainty rather than regions that improve the regression

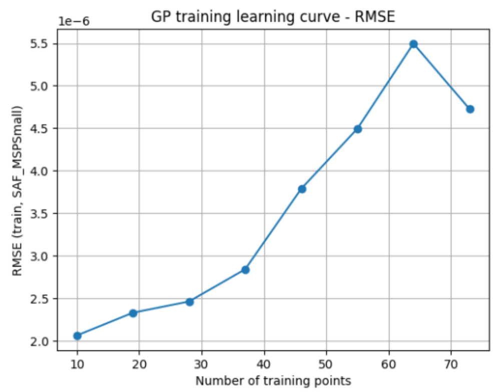

Figure 5: GP Training Learning Curve – RMSE

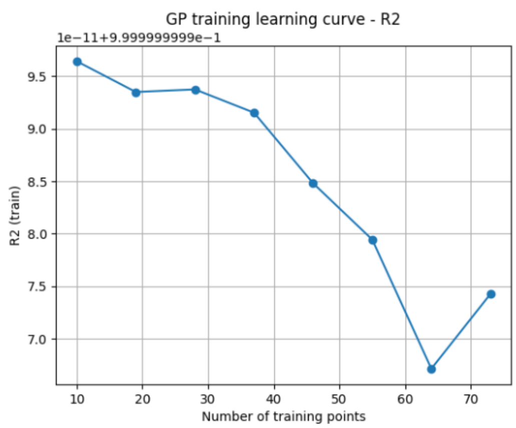

Figure 6: GP Training Learning Curve$-R^2$

fit. As a result, the observed decrease in training R^2 is consistent with a model that is progressively expanding coverage of the input space and reducing epistemic uncertainty rather than optimizing predictive accuracy on the training set.

5.2 Bayesian Optimization

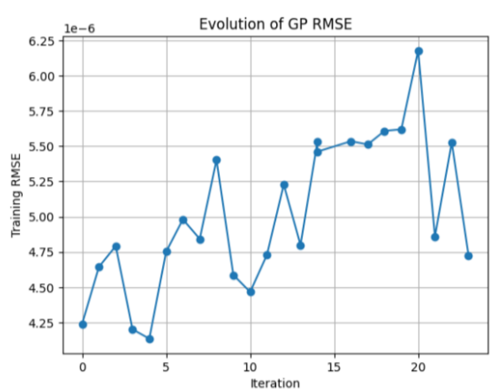

Figure 7: Training RMSE vs. iteration

The training RMSE exhibits fluctuations and a gradual upward trend as iterations progress. This behavior is characteristic of exploration-driven Bayesian optimization, where the acquisition function deliberately selects points in regions of high uncertainty rather than those that minimize prediction error. As a result, the GP surrogate must fit increasingly diverse and challenging samples, which naturally increases the training residual. This indicates that the optimization process is expanding its coverage of the design space rather than refining the fit around previously sampled areas.

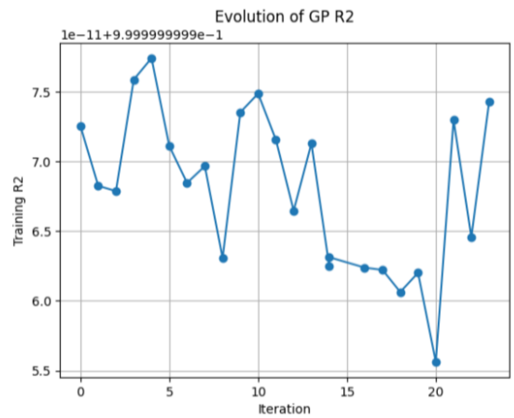

Figure 8: Training R^2 vs. iteration

The R^2 metric decreases over the course of the optimization, with oscillations reflecting the non-uniform difficulty of the sampled points. This reduction in goodness-of-fit is expected during exploratory BO: as new points are added in poorly understood regions, the GP must compromise its earlier interpolation accuracy to represent a broader function space. The temporary R^2 drops, especially around mid-to-late iterations, confirm that the algorithm is prioritizing uncertainty reduction over immediate predictive accuracy.

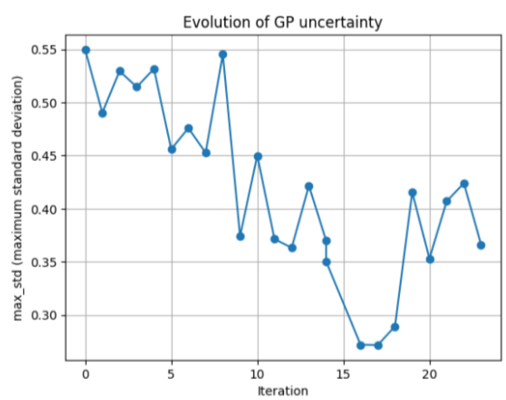

Figure 9: Maximum uncertainty (max_std) vs. iteration

Unlike RMSE and R^2, the evolution of maximum predictive standard deviation shows a clear downward trend. The progressive reduction in max_std demonstrates that Bayesian optimization is successfully decreasing epistemic uncertainty across the domain. Despite local oscillations, caused by alternating exploration in different regions, the general decline indicates that each new sample contributes to reducing the GP's uncertainty envelope. This is a key indicator that the acquisition strategy is effective and that the surrogate model is becoming more globally informed.

6 Conclusion

This study demonstrates the effectiveness of combining Gaussian Process (GP) surrogate modeling with Bayesian Optimization (BO) to evaluate and minimize the Minimum Selling Price (MSP) of Sustainable Aviation Fuel (SAF) produced via the Alcohol-to-Jet (ATJ) pathway. Using a System Dynamics model developed in Stella Architect, we generated techno-economic data across eight key economic drivers and showed that a GP surrogate with an ARD–RBF kernel can accurately reproduce the nonlinear behavior of the ATJ system. The surrogate achieved near-perfect interpolation of the simulation data and successfully identified ethanol price as the dominant cost driver influencing MSP.

The integration of GP modeling with an uncertainty-driven acquisition function enabled efficient exploration of the economic design space. The BO process systematically reduced epistemic uncertainty, expanded coverage into unexplored regions, and revealed a minimum MSP of 0.318 USD/L within the evaluated domain. Learning-curve trends further highlighted the expected trade-off between exploration and predictive accuracy, with training error increasing as the model incorporated samples from high-uncertainty regions. These results collectively demonstrate the utility of surrogate-assisted optimization as a scalable and computationally efficient approach for techno-economic assessment. This promises a positive social impact as this analysis could lead to increased economic efficiency and the promotion of efficient approaches for techno-economic assessment.

A key limitation arises from the absence of a direct interface between Stella Architect and Python, which required manual data transfer and constrained the total number of optimization iterations. Automating this process would enable larger experimental designs, stronger generalization of the GP surrogate, and more effective global optimization. Future work should incorporate automated simulation pipelines, alternative acquisition strategies, and multi-output GP models to capture a broader range of techno-economic indicators. Despite these practical constraints, the methodology presented here highlights a promising direction for integrating simulation and machine learning to accelerate SAF cost analysis and inform strategic decision-making.

7 Supplementary Material

Gaussian_Process_For_MSP_of_Sustainable_Aviation_Fue.ipynb

570_Gaussian_Process.git

References

E. Badakhshan et al. Application of simulation and machine learning in supply chain management. *Computers & Industrial Engineering*, 198:110649, 2024. doi: 10.1016/j.cie.2024.110649.

K. Brandt and M. Wolcott. Cumulative impact of policy on the minimum selling price of sustainable aviation fuel. *Frontiers in Energy Research*, 10:884947, 2022. doi: 10.3389/fenrg.2022.884947.

A. Damianou, M. Titsias, and N. Lawrence. Variational gaussian process dynamical systems. In *Advances in Neural Information Processing Systems*, volume 24, pages 2510–2518, 2011. PDF analyzed in this study.

Paulina Echeverria Paredes. System dynamics model of the atj techno-economic system. PhD student in Biological Systems Engineering, Washington State University. Model developed in Stella Architect. Not yet published., 2025.

S. Geleynse et al. The alcohol-to-jet conversion pathway for drop-in biofuels: A review. *ChemSusChem*, 11(21):3728–3741, 2018. doi: 10.1002/cssc.201701589.

isee systems, inc. Stella architect (software). Software product page, 2025. URL `https://www.iseesystems.com/store/products/stella-architect.aspx`.

J. J. Lucas, R. A. Ruiz-Mercado, M. M. Andersen, and D. E. Cresswell. Surrogate modeling strategies for chemical process simulation: A review. *Computers & Chemical Engineering*, 151:107343, 2021. doi: 10.1016/j.compchemeng.2021.107343.

B. Meyer et al. Using gaussian process regression for surrogate modeling in chemical engineering: A review. *Chemical Engineering Science*, 234:116318, 2021. doi: 10.1016/j.ces.2020.116318.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006. doi: 10.7551/mitpress/3206.001.0001.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1): 148–175, 2016. doi: 10.1109/JPROC.2015.2494218.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

John Sterman. System dynamics: systems thinking and modeling for a complex world. 2002.

Alternative Fuels Data Center U.S. Department of Energy. Sustainable aviation fuel. `https://afdc.energy.gov/fuels/sustainable-aviation-fuel`. Accessed: 2025-12-07.

L. von Rueden et al. Combining machine learning and simulation to a hybrid modelling approach. In *Lecture Notes in Computer Science*, volume 12080, pages 548–560, 2020. doi: 10.1007/978-3-030-44584-3_43.

J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models. In *Advances in Neural Information Processing Systems*, volume 18, pages 1441–1448, 2005. PDF analyzed in this study.

G. Yao et al. Stochastic techno-economic analysis of alcohol-to-jet fuel production. *Biotechnology for Biofuels*, 10(244), 2017. doi: 10.1186/s13068-017-0840-5.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction include the claims made in the paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations section discusses this topic.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Yes, the theoretical framework, data, and model are all present.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Yes, the paper documents our full methodology, theoretically and including code.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The paper includes access to the time series data and code in the github repository.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Yes, the methodology and available code documents all such details.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper does not include experiments that require an analysis of statistical significance.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The methodology is not computationally expensive such that special compute resources are required.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the code of ethics and find our work compliant.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper discusses its potentially positive social impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper does not pose such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All creators and owners of assets are cited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: New assets, such as the model and data, are available on the corresponding github repository that is linked in the supplementary material section.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects.**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs are not involved as any important, original, or non-standard component of our methodology.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.