PURDUE UNIVERSITY | The Data Mine

STRATOLAUNCH

Achieving Hypersonic Similitude

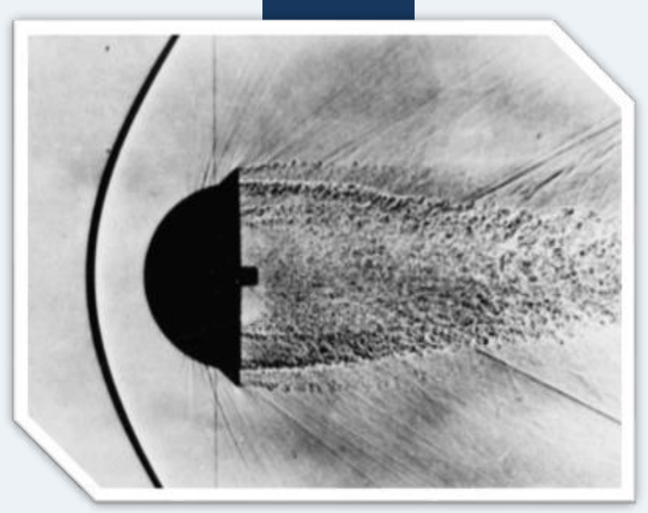
Project Team: *Benjamin Dravis, Kalyan Niranjan Archakam, Ekansh Agrawal, Owen Willis, Ethan Lee, Andrew Kachur, Benjamin Stransky, Kevin Ma, Aditya Singh*

INTRODUCTION

Our aim is to analyze ***hypersonic wind tunnel data*** from physical testing to make ***computer-generated CFD*** data more representative of test data. This will enable CFD to play a larger role in vehicle design, and lead to quicker development cycles and a better product.

We are developing an app that uses statistical analysis to detect and eliminate noise in wind tunnel simulations. Our goal is to calibrate computational models with wind tunnel test results that are more accurate but more difficult to conduct. The app will use graphing, statistical modeling, and LLM output to simplify the process of understanding the wind-tunnel data. We will use insights from aerodynamic theory to interpret the results.

Boeing Mach-6 Quiet Tunnel (BAM6QT) (1)

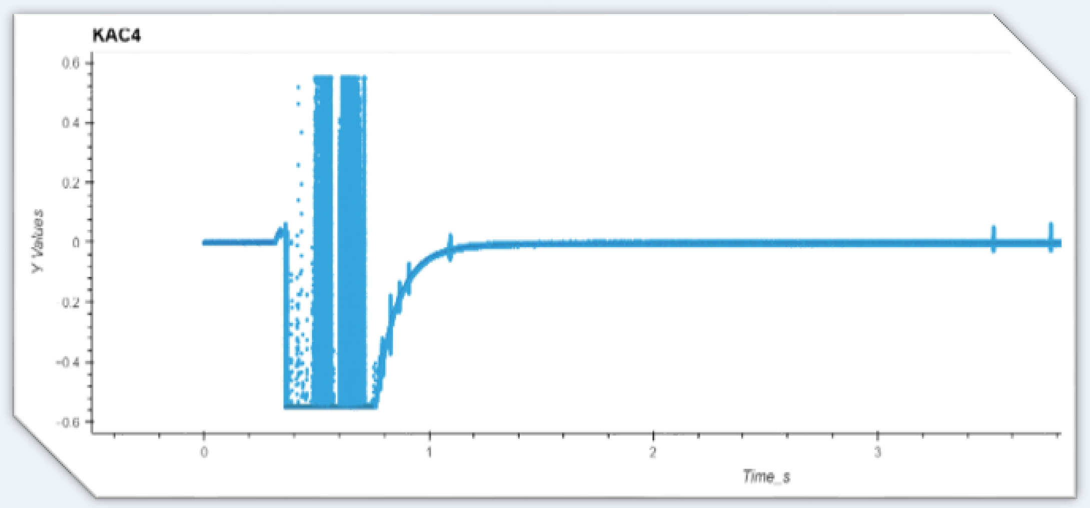


Sample Bow Shock Experienced at Nose (2)

LITERATURE REVIEW

The Data Interpretation Team studied wind tunnel research produced at Purdue, and the standard academic texts on hypersonic flow. *Hypersonic Aerothermodynamics* by John Bertin was the main source of information on general hypersonic characteristics. To learn about the BAM6Q wind tunnel, the team studied papers from Dr. Joseph Jewell and Dr. Steven Schneider. From these, the team was able to identify the primary sources of noise in the BAM6Q tunnel.

Our study enabled us to identify/categorize tunnel noise, locate data disruptions and associated causes, and determine methods of modeling this noise for comparison with CFD.



Pressure Sensor Data from Wind Tunnel (5)

Project Goal

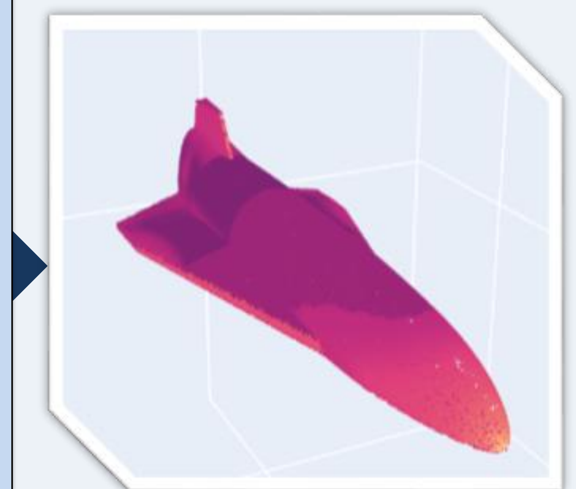
- Improve Stratolaunch's CFD accuracy by comparing simulations with wind tunnel data.
- Develop methods to identify discrepancies in:
  - Transition points
  - Disturbed flow regions
  - Shockwaves (Figure 2)
- Analyze differences to uncover CFD error sources and support model refinement.
- Enhance the predictive accuracy of CFD for hypersonic applications

Computational Fluid Dynamics (CFD) Data

- Provides a steady-state snapshot of fluid behavior.
- Generates detailed 3D X-Y-Z data for all variables.
- Enables full-vehicle visualization of aerodynamic and thermal properties.
- Based on idealized simulations, so real-world accuracy is limited
- Figure (3): Temperature distribution across Stratolaunch vehicle highlights spatial variation.
- Real-world comparison is needed to improve CFD reliability and model fidelity.

Wind Tunnel Data

- Captures real-world, time-dependent aerodynamic behavior.
- Uses discrete sensors to collect data, but coverage is spatially limited and noisy.
- Time-series includes:
  - Transient and steady-state components
  - Requires filtering and careful time slice selection for CFD comparisons.
- Figure (5): Data from BAM6QT facility tests shows key performance trends.
- Essential for identifying:
  - Short-term fluctuations
  - Steady behavior
  - Unique flow phenomena
- Provides experimental validation, critical to refining CFD models.

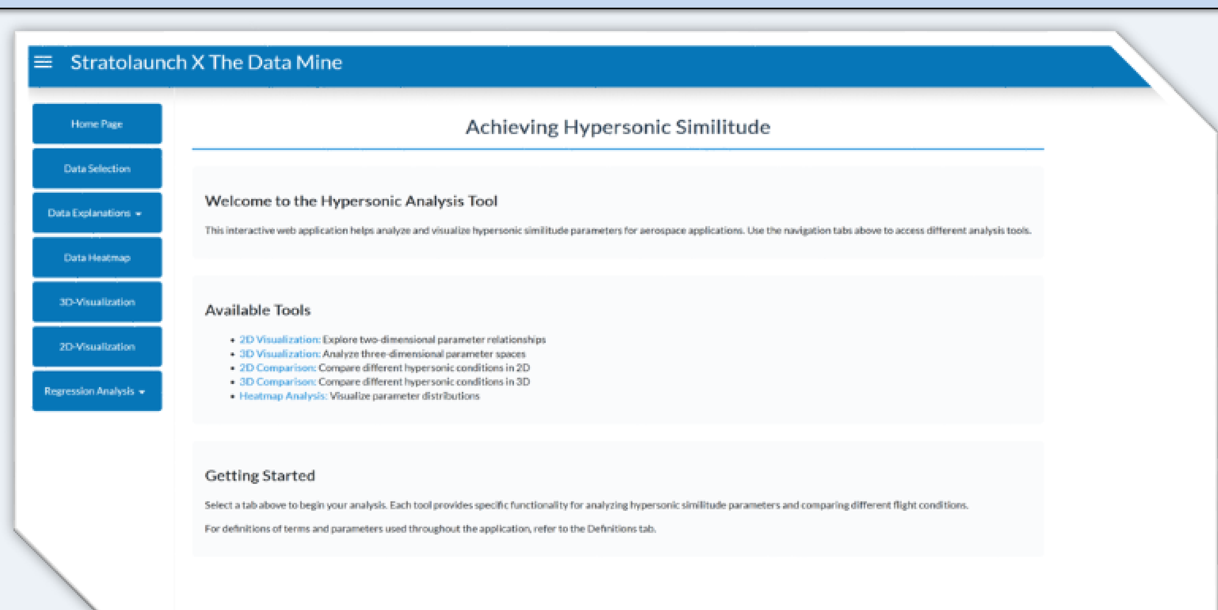


CFD Aircraft Visualization (3)

Aircraft Used in Wind Tunnel (4)

PANEL APPLICATION

- We developed a web application using Python as an interactive data analysis platform for this project, and to accomplish our goals featuring several different specialized tabs:
  - Data viewing for looking through a data sets and their columns
  - Viewing heat map correlations to determine what would be best to visualize with the other plotting tools of the application
  - 2D/3D visualizations both for singular and comparison-based plots to look at data in a data set or compare it against data within the same set or a different set
  - Advanced statistical modeling for data predictions:
    - Linear/polynomial regression and corresponding plots
    - random forest algorithms and corresponding plots



Application Homepage

LARGE LANGUAGE MODELS (LLM)

Our project uses open-source Large Language Models (LLMs) to make data analysis easier and smarter. By integrating different types of LLMs, our app helps users better understand and work with complex datasets:

• **General LLMs** (e.g., Llama 2) explain trends and patterns in plain language.

• **Code-focused LLMs** (e.g., StarCoder, Code Llama) improve algorithm analysis and code-based outputs.

• **Math & Science LLMs** (e.g., GPT-NEOX) boost statistical analysis and predictions.

With these models, users can:

• Summarize data and graphs
• Analyze trends and regressions
• Translate technical info into clear insights

Overall, we make data interpretation more intuitive and accessible.

CONCLUSION

Our analysis of Ollama models shows that larger models (32B+) generate more coherent, contextually relevant insights but at the cost of higher computation and latency. Smaller models are faster and more efficient but hallucinate more, especially with weak retrieval, struggling with nuanced reasoning. Balancing model size, retrieval quality, and efficiency is key to integrating LLM-driven insights into structured data workflows.

Additionally, to improve the accuracy of CFD data, our process was to introduce artificial noise to better match wind tunnel data and analyze how the transition location shifts due to it. Wind tunnel tests tend to operate at different Reynolds numbers than CFD data, which often affects this point and aerodynamic stability. Using dimensionless parameters helped correct these differences within the Reynolds number and theoretically would help better align CFD and wind tunnel results.

NEXT STEPS

For LLM next steps, we'll enhance retrieval accuracy, reduce hallucinations in smaller models, and optimize efficiency. Key steps include refining custom retrieval for high-quality responses, experimenting with model distillation for smaller models, and exploring hybrid retrieval (e.g., vector search + symbolic reasoning).

For aerospace predictions next steps, we plan on incorporating artificial noise into CFD models and momentum thickness-based Reynolds numbers for more accurate comparisons. Applying these conditions to flight conditions will improve the reliability of Stratolaunch's hypersonic CFD data.

ACKNOWLEDGEMENTS

Thank you to Dr. Steven Hunt, Conrad McGreal, Dr. Ward, Kevin Amstutz, Kevin Boes, Rohit Ramkumar, Dr. Joseph Jewell, Samantha Miller

The Data Mine Corporate Partners –> Symposium 2025