The Ariel Data Challenge: Crowdsourcing reveals the limits of deep learning for Space Exploration

K. H. Yip, R. L. Coates,
L. V. Mugnai, A. Papageorgiou,
A. Bocchieri, O. Faucoz, A. Syty,
T. Tahseen, I. Waldmann,
& the Ariel Consortium

Dr. Kai Hou (Gordon) Yip
kai_hou.yip@kcl.ac.uk

1. ARIEL DATA CHALLENGE

Ariel is a UK-led ESA space telescope mission that will survey the atmospheres of ~1000 exoplanets (planets that orbit stars other than our own sun), over a period of 4-6 years, due to launch in 2031.

The Ariel Data Challenge is a global online competition that invites the machine learning community to tackle difficult problems in the cleaning and analysis of space telescope data when surveying exoplanets.

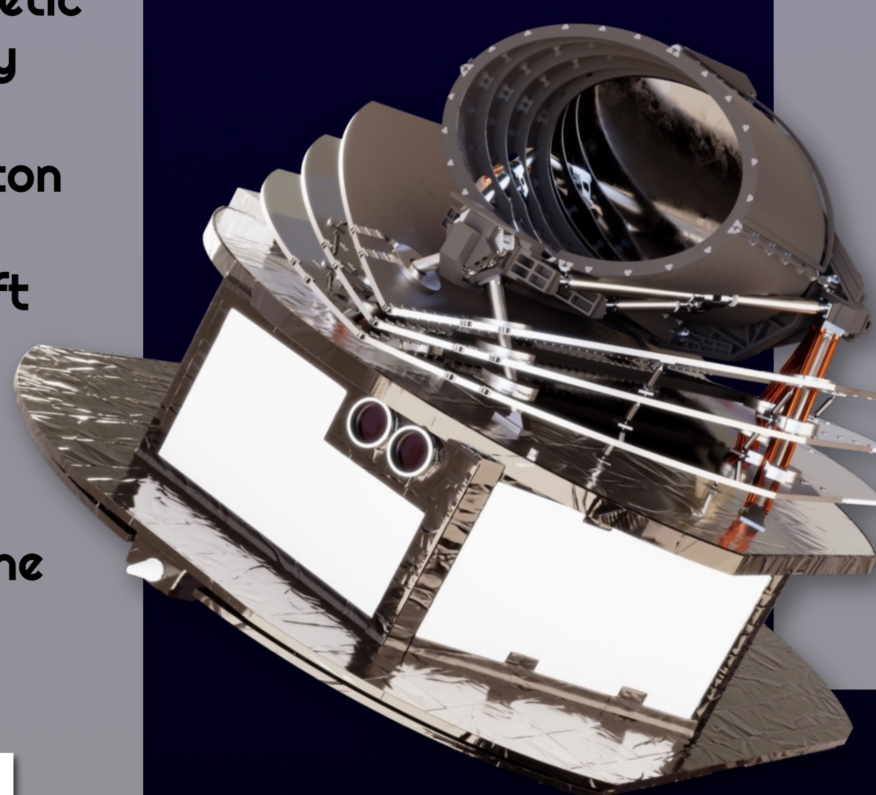
Participants compete over three months to provide optimal solutions to a different exoplanet data problem each year. The resulting insights are integrated back into Ariel science, enhancing mission capabilities and scientific yield.

2. METHODS

We simulate the mission's data with the use of official Ariel instrument simulator, ExoSim 2. Two out of six Ariel instruments were simulated, i.e. FGS-1 (0.6 – 0.8 μm) and AIRS-CH0 (1.95 – 3.90 μm). To ensure our data is as realistic as possible, we implemented:

1. Diverse Atmospheric profiles – we simulated different atmospheric models for both known and unknown planets to highlight the diversity in atmospheric profiles and ensure solutions are scientifically useful.
2. Realistic Ariel Targets – we generated synthetic planets from actual distribution of planetary parameters (Rp, Mp, Tp etc).
3. Realistic Instrument noise – we injected photon noise, gain drift, jitter effects and actual calibration frames from the latest spacecraft (JWST NIRSpec).

A total of 2400 simulations were created. Half are used as training data for the participants, and the other half used as a hidden test set. The number of training data is reference to the expected yield of Ariel.



3. EVALUATION PROTOCOL

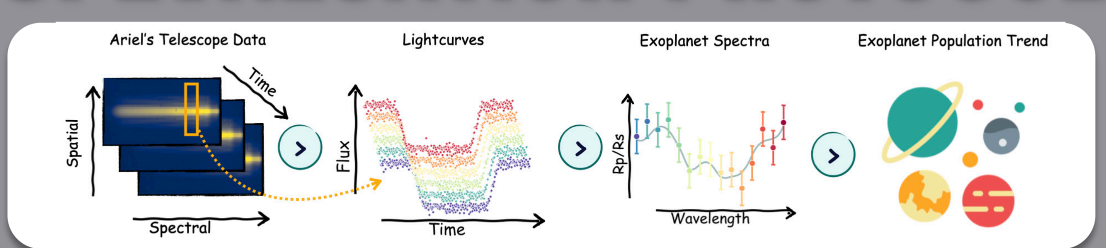


Figure 1: An illustration of Ariel's data product and science goals.

Each team submits a spectroscopic measurement of each target (Ariel's science product) to be graded by comparing their differences with the ground truth using a scaled log-Gaussian Likelihood rubric. The score spans 0 (no match) to 1 (a perfect match.) The ideal performance of Ariel would achieve a score of ~0.9.

4. RESULTS

I. **Participation.** 1,000+ teams from 75+ countries, averaging 260 submissions per day over 3 months (~25,000+ total). Demographics split roughly equally across academia, students, and industry professionals.

II. **Simple, physics-informed models dominate.** Meta-analysis across all submissions shows that simple ensemble methods with physics-informed features, grounded in instrument and atmospheric understanding, consistently outperform complex deep learning approaches. The top 10 solutions were dominated by these interpretable methods, scoring far above the median of 0.33.

III. **Bayesian inference validated as optimal for the mission.** Bayesian techniques ranked among the top solutions across both editions, taking 1st place in 2025. The winning innovation combined multidimensional GPs to model stellar and noise components with fast optimisation, independently verifying the mission's chosen approach.

IV. **Deep learning fails on out-of-distribution data.** The evaluation metric filters out underperforming approaches, and deep learning models were systematically eliminated. This directly contradicts standard machine learning literature in astronomy, where deep learning is prevalent.

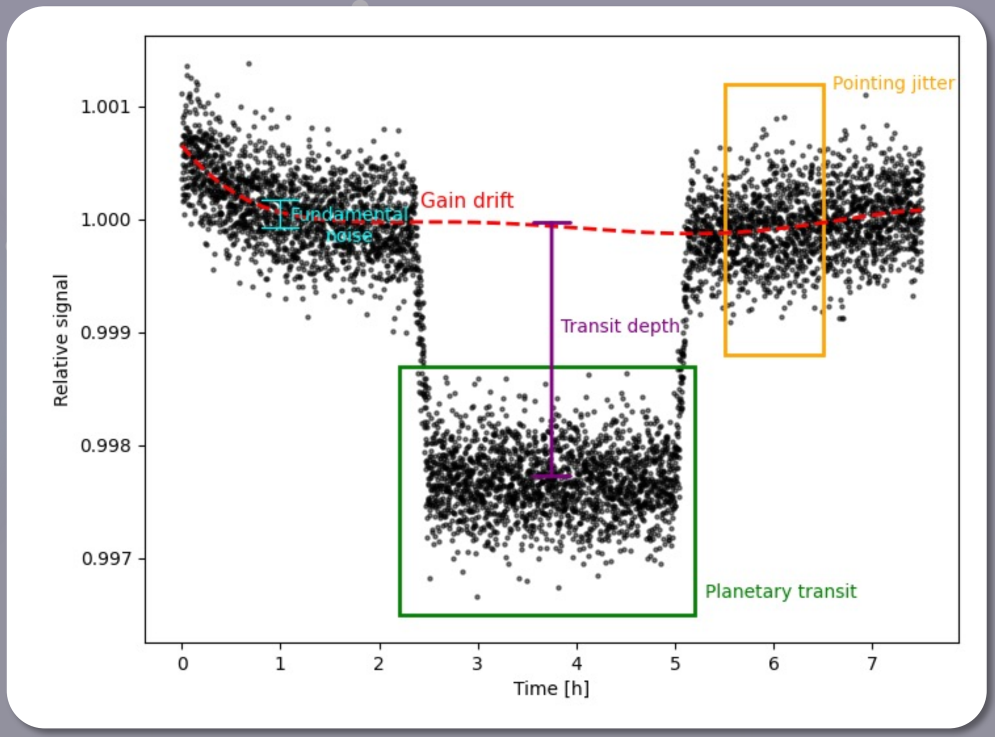


Figure 2: Schematic of different noise components on a light curve of a transiting exoplanet.

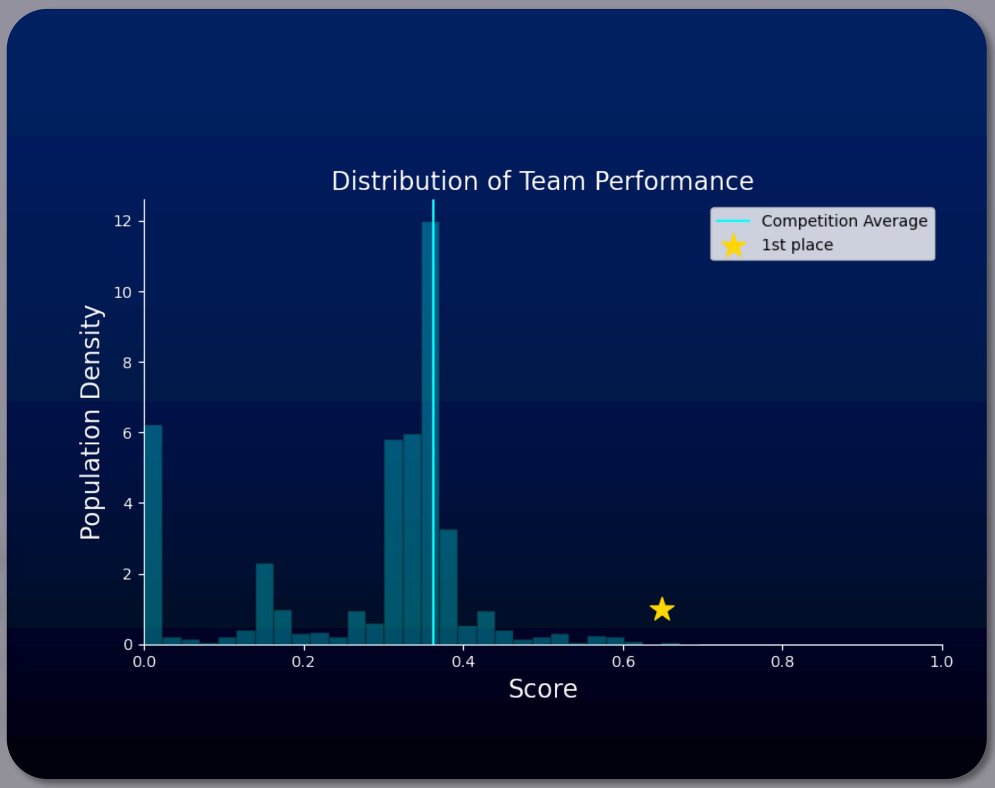


Figure 3: Distribution of scores of different teams in the competition. The Star highlights the first place.

5. IMPACT

The Ariel Data Challenge is the first major attempt to integrate exoplanet science with wider fields such as Machine Learning. It is known for its physical fidelity grounded in highly realistic physics simulation.

The challenge has successfully cultivated international recognition and financial support of more than USD $100,000 from sources including STFC Scientific Computing, CNES (French Space Agency), and Kaggle (Google's data science platform), who also provided the platform on which to hold the competition for 2024 and 2025.

Credits

cnes · iAP · UCL · DiRAC · Cardiff University · Sapienza Università di Roma