NEURAL INFORMATION PROCESSING SYSTEMS

kaggle

7th Place Solution: A Hybrid Physics-Based Method with Neural and Tree-Based Model Corrections

YuXuan Wu* & Katio Takano

Ariel DATA CHALLENGE

Background

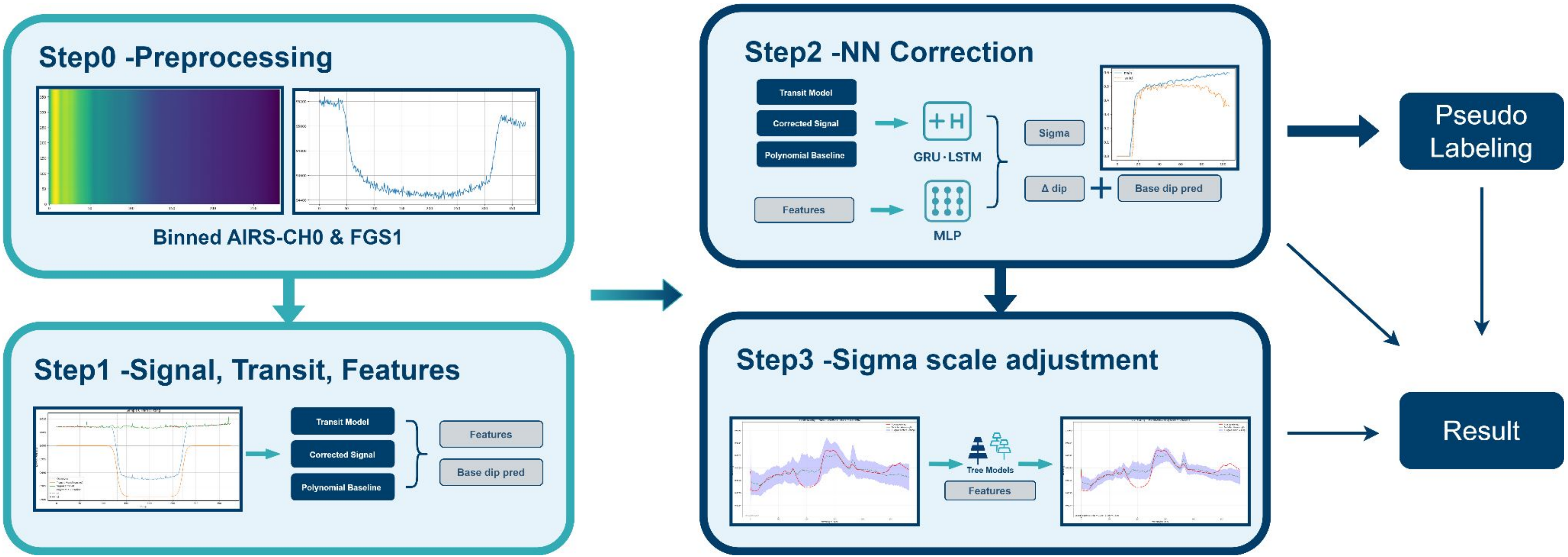
- YuXuan Wu (Kaggle ID: Horikita Saku）
  - Research Assistant at *China National Center for Bioinformation*
  - Machine Learning / Genetics / Single-cell omics
  - Interested in Astronomy
  - Applying for PhD. horikitasaku@outlook.com

- Kaito Takano (Kaggle ID: takaito）
  - Quantitative Analyst at *Nomura Asset Management Co., Ltd.*
  - Visiting Researcher at *Osaka Metropolitan University*
  - Finance / Natural Language Processing / Machine Learning
  - Ph.D. (Science and Technology)
  - I like competitions!

Agenda

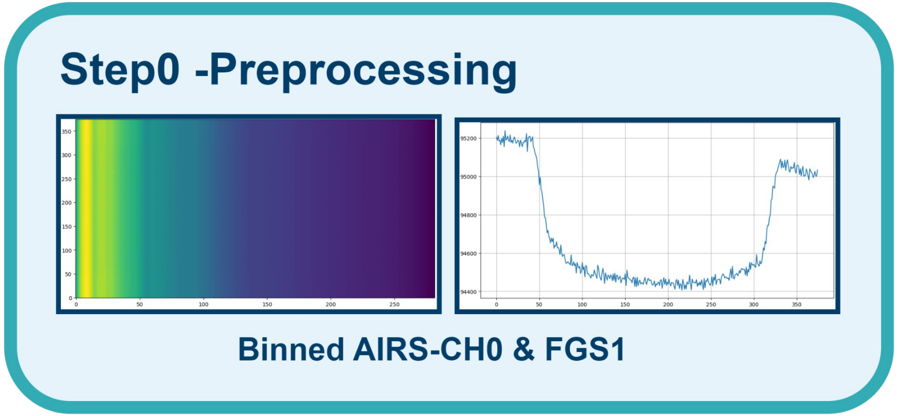
1. Background
2. **Main Solution**
   - Step0: Preprocessing
   - Step1: Signal, transit, and feature extraction
   - Step2: Neural Network Correction
   - Step3: Sigma scale adjustment
   - Step4: Pseudo Labeling
3. What didn’t work
4. Summary

Main Solution : Overview

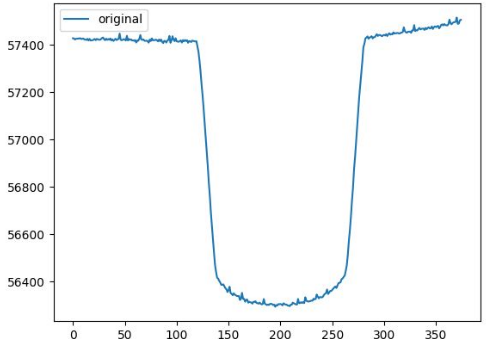


Step0: Preprocessing

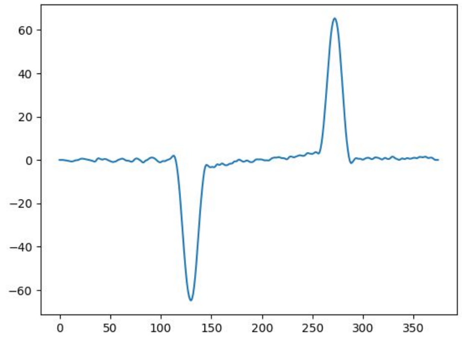
By *Ariel 2025*, we failed to discover any more magic or unique perspectives. Therefore, allow me to skip this part.



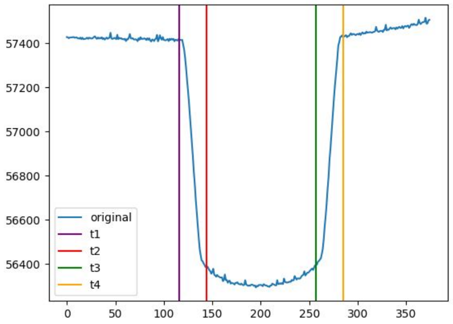
Step1.1: Phase Detector



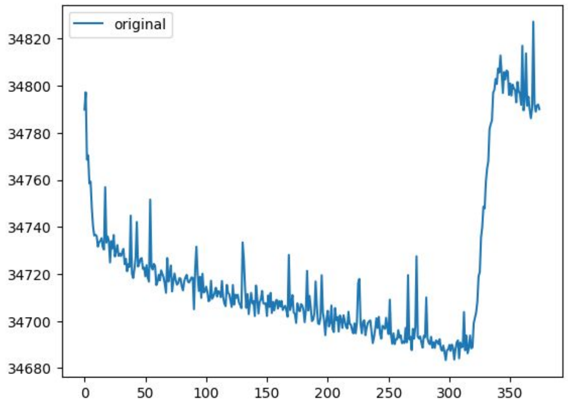
Smoothed Averaged Raw signal



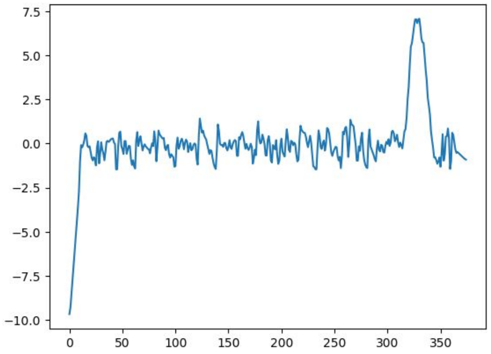
Smoothed Gradient

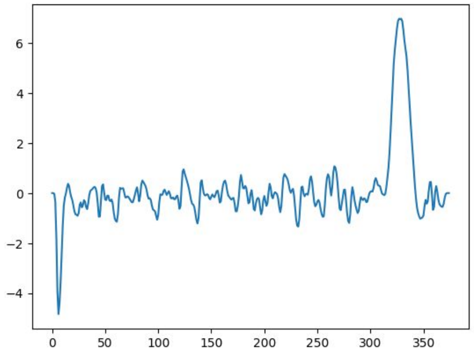


Step1.1: Phase Detector - Robustness



Planet 561423413





Step1.1: Phase Detector - Robustness

50 Samples From pre-ingress

50 Samples From ingress-transit

Difference

depth_samples_ingress

$$\mu_{\text{in}} = \mathbb{E}\big[\text{depth\_samples\_in}\big],$$

$$\sigma_{\text{in}} = \text{Std}\big[\text{depth\_samples\_in}\big],$$

depth_samples_egress

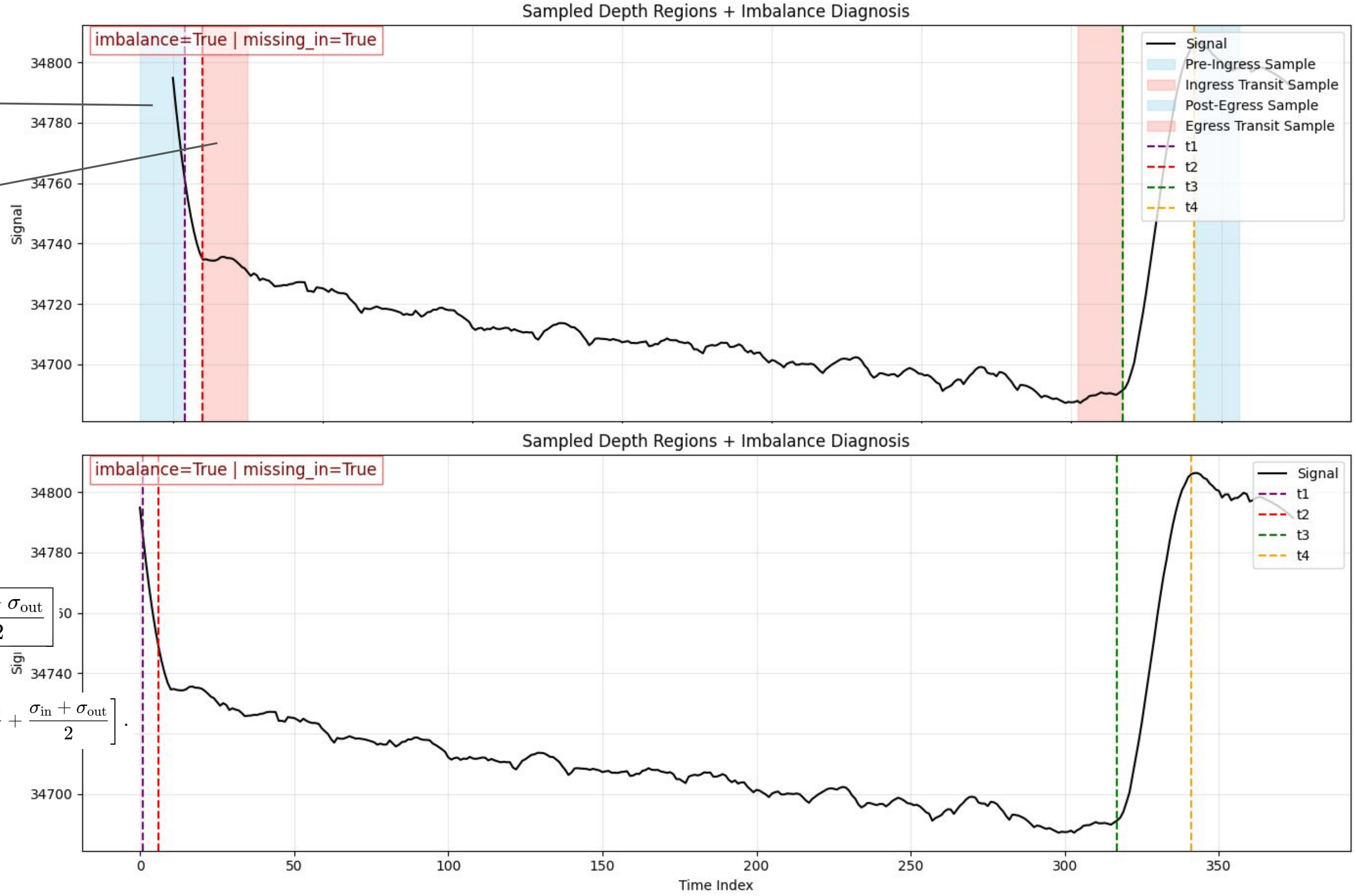
$$\mu_{\text{out}} = \mathbb{E}\big[\text{depth\_samples\_e}\big],$$

$$\sigma_{\text{out}} = \text{Std}\big[\text{depth\_samples\_e}\big].$$

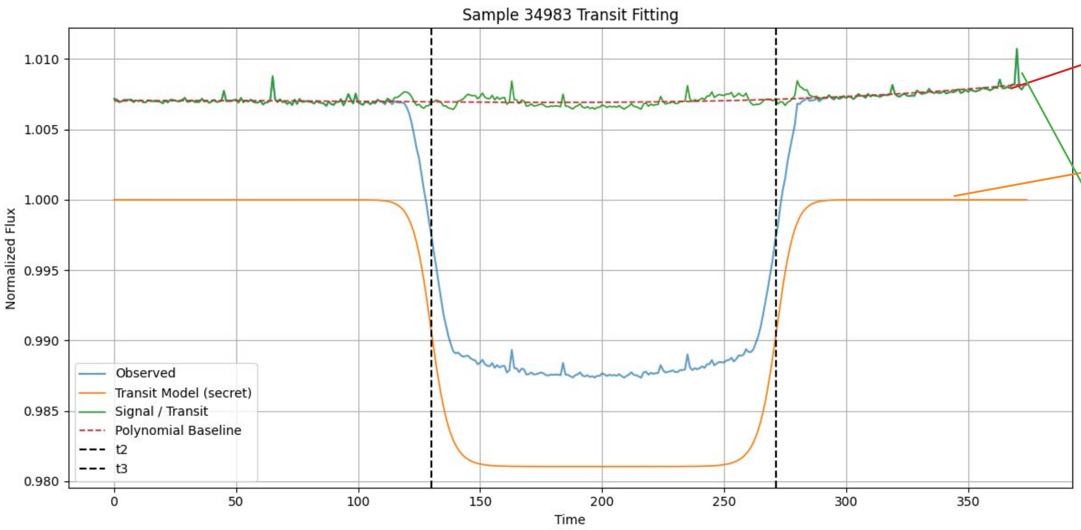
$$\boxed{\text{depth\_gap} = |\mu_{\text{in}} - \mu_{\text{out}}| + \frac{\sigma_{\text{in}} + \sigma_{\text{out}}}{2}}$$

$$\text{imbalance} = \left[\text{depth\_gap} > \frac{\mu_{\text{in}} + \mu_{\text{out}}}{2} + \frac{\sigma_{\text{in}} + \sigma_{\text{out}}}{2}\right].$$

Check whether the "depth_gap" is greater than the "upper bound of the normal difference"



Step1.2: Physics-based modeling and feature extraction



Polynomial Baseline

Obtained from the OOT

Following the approach of 2024 10th, define a bell-shaped function as the ideal model.

Fitting

Reconstructed signal:
If there is no transit

SOLUTION WRITEUP

10th Place Solution

Thanks to the participants and organizers for this competition. It was very interesting and educational. I hope to see Ariel on kaggle in a year. Approach We used a polynomial approximation approach. Thanks a lot to...

Nov 4, 2024 · NeurIPS - Ariel Data Challenge 2024 · 10th Place

XoMoX

...

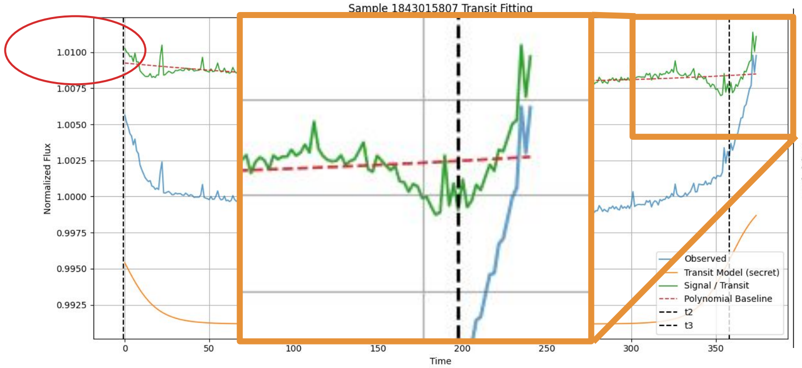
TALK
IN
COMPETITION: ARIEL DATA CHALLENGE 2024: EXTRACTING EXOPLANETARY SIGNALS FROM THE ARIEL SPACE TELESCOPE

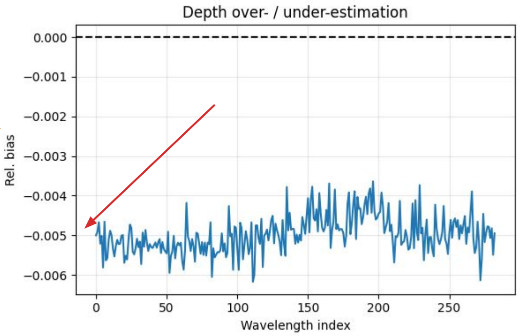
10th place: polynomial approximation, does it solve the problem?

Georgii Aparin

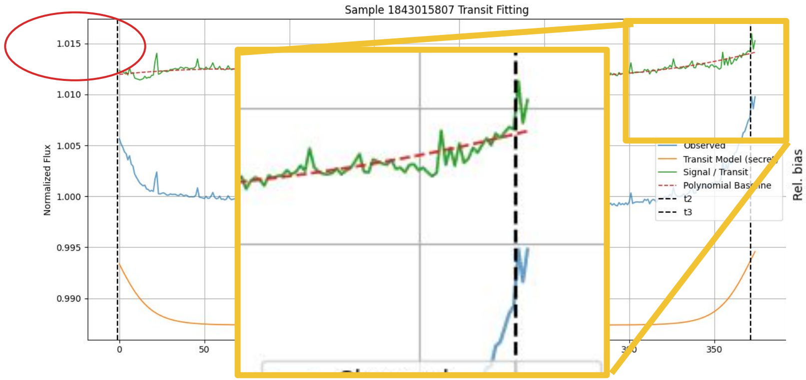
2024 Talk
in
Competition: Ariel Data Challenge 2024: Extracting exoplanetary signals from the Ariel Space Telescope

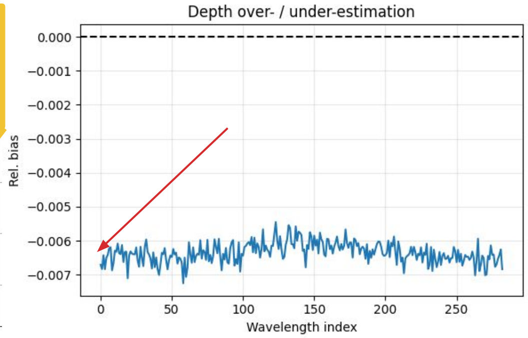
Step1.2: Low-degree polynomials and GP





Degree 2

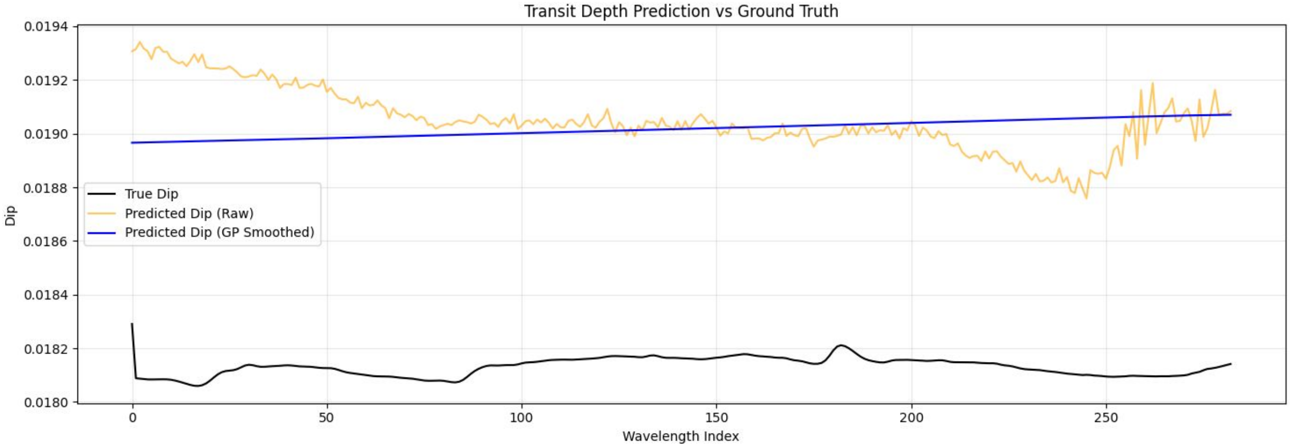




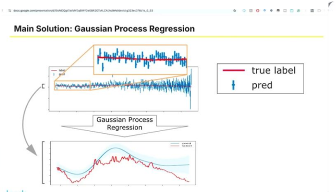
Degree 3

⭐The complexity of the baseline, and the OOT used for fitting the baseline

Step1.2: Low-degree polynomials and GP



Utilize GP to smooth the signal.
It works for all downstream corrections!



TALK
IN
COMPETITION: ARIEL DATA CHALLENGE 2024: EXTRACTING EXOPLANETARY SIGNALS FROM THE ARIEL SPACE TELESCOPE

1st Place Solution: Leveraging Knowledge of the Physical Model

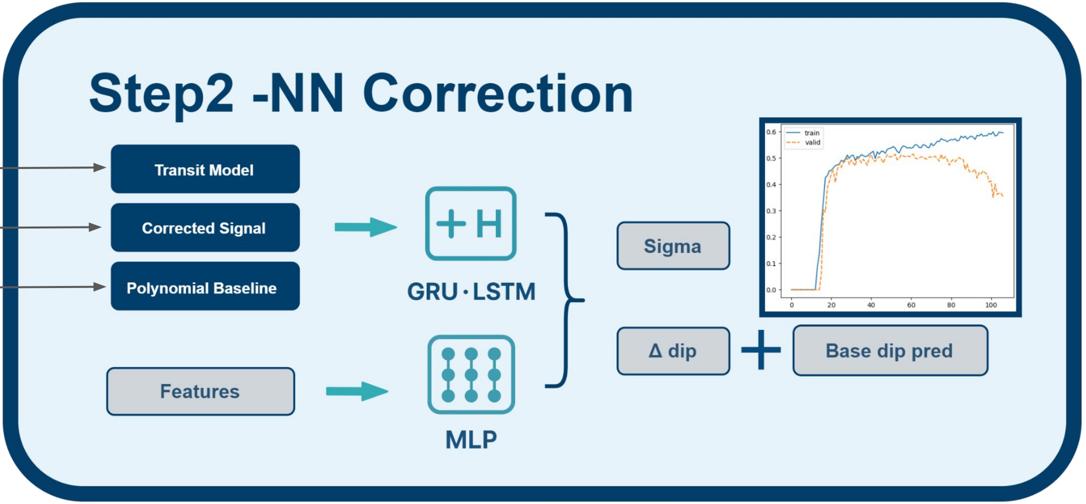
Kohki Horie · Yamato Arai

2024 Talk
in
Competition: Ariel Data Challenge 2024: Extracting exoplanetary signals from the Ariel Space Telescope

Step2: Difference correction with NN and base sigma prediction

Different combinations

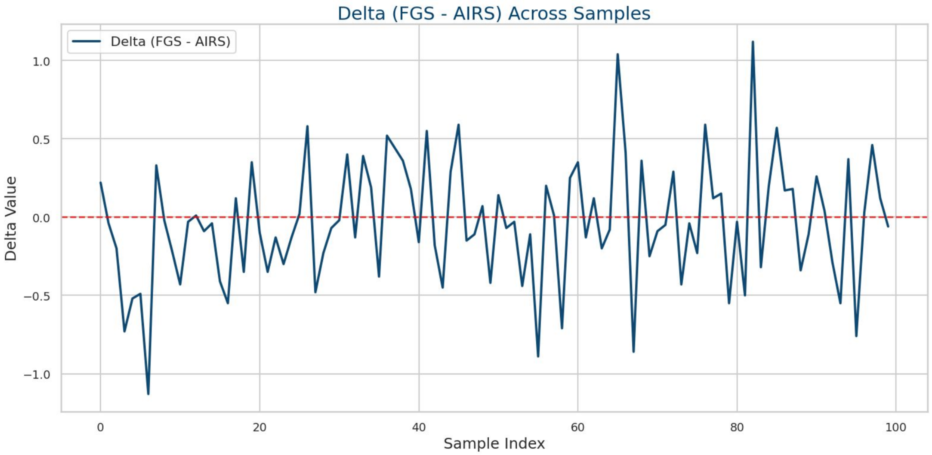
- Multilayer Perceptron (MLP)
- BiGRU + CNN + MLP
- BiLSTM + CNN + MLP
- BiLSTM + CNN + BiLSTM



Key techniques

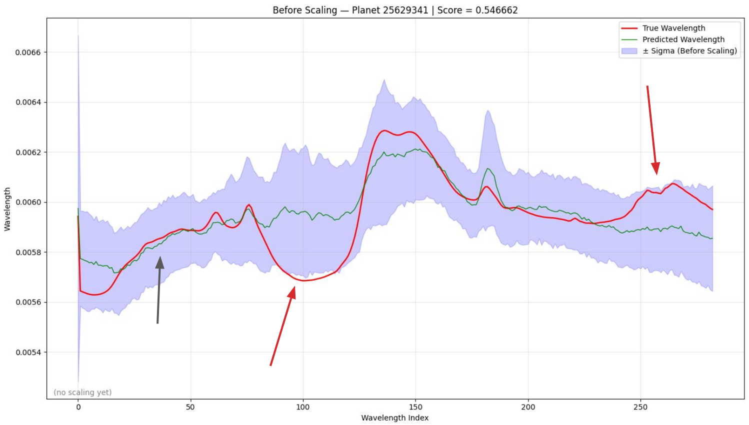
- Used **quantile regression** so the model can also predict sigma.
- Applied **Adversarial Weight Perturbation (AWP)**.
- Performed data augmentation by flipping signal and transit sequences along the time axis.
- Added noise to the data for further augmentation.

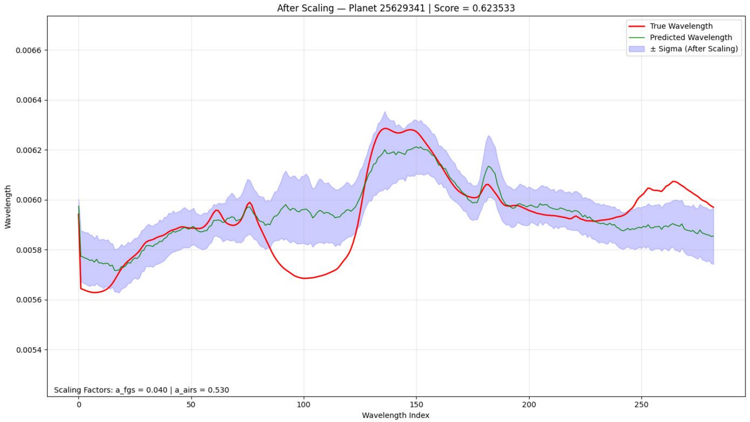
Step3: Sigma scale adjustment with Gradient Boosting



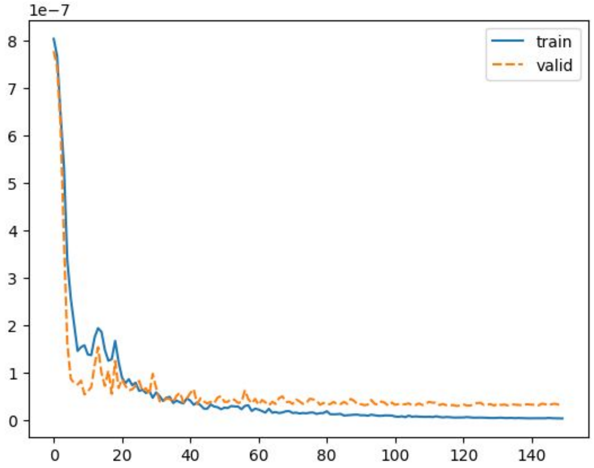
Separate the scaling factors for FGS and AIRS.
a_fgs * σ FGS : Scaling factor for FGS
a_airs * σ AIRS : Scaling factor for AIRS-CH0

Learned by LGBM/CatBoost/XGB

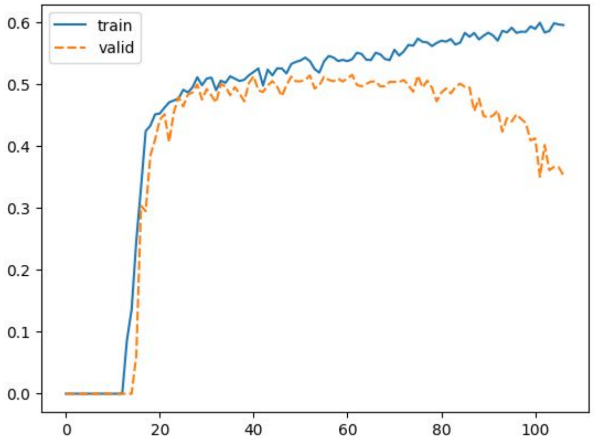




Step4: Pseudo Labeling



MSE transition
Almost no folds showed clear overfitting.



Competition metric
Tended to overfit more easily.

we only used the predictions for **dip**

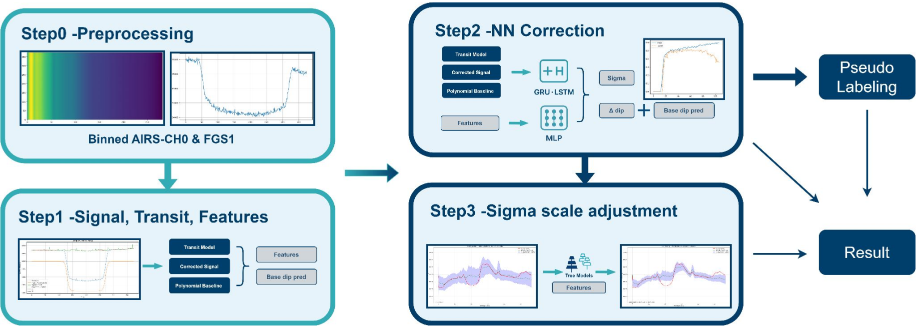
What didn’t work

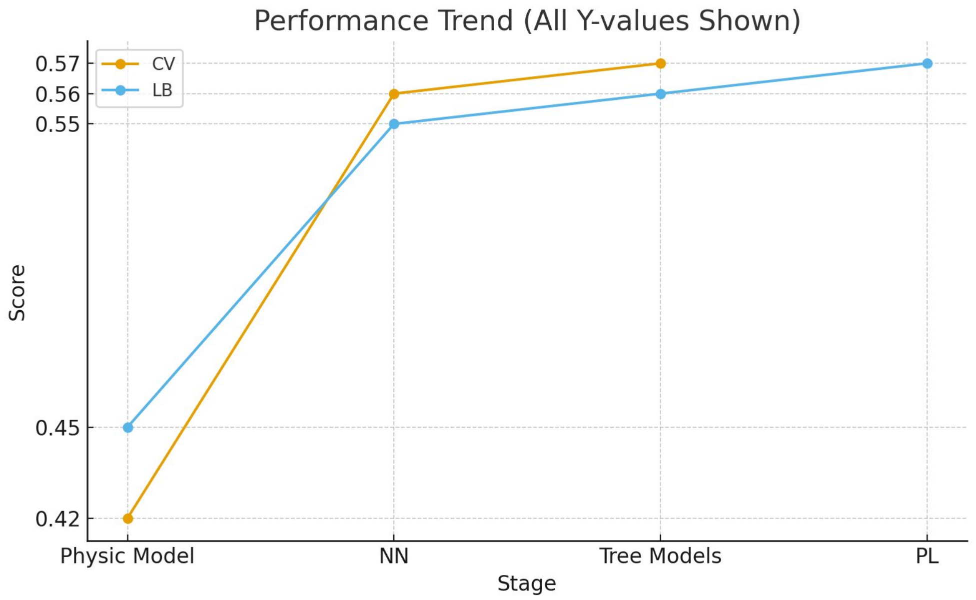
1. More Complex Physic Models
2. Using 1st Place Solution from ADC 2024’s Physic Model
3. More Complex NN model architectures
4. Some Magics during preprocessing

Agenda

1. Background
2. Main Solution
    - Step0: Preprocessing
    - Step1: Signal, transit, and feature extraction
    - Step2: Neural Network Correction
    - Step3: Sigma scale adjustment
    - Step4: Pseudo Labeling
3. What didn’t work
4. **Summary**

Summary





The scores at each stage are highly correlated with each other.
By using physical models to assist neural networks, we can enhance the **generalization** ability and **robustness** in different situations.

Thanks!

I'm currently seeking a PhD opportunity. Please contact me!

horikitasaku@outlook.com