Correcting Transiting Exoplanet Light Curves for Stellar Spots

Discovery Challenge

ARIEL Science, Mission & Community Conference 2020

ARIEL
europlanet SOCIETY
UK SPACE AGENCY
Horizon 2020
erc European Research Council Established by the European Commission
UCL CENTRE FOR SPACE EXOPLANET DATA

I’m not Nikos Nikolaou

This is Nikos Nikolaou

The Data Challenge Team

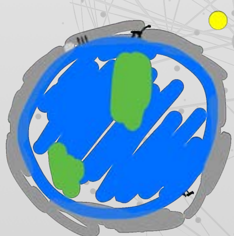
Nikos Nicolaou
Project lead

Angelos Tsiaras
Data set generation

Subi Sarkar
Data set generation

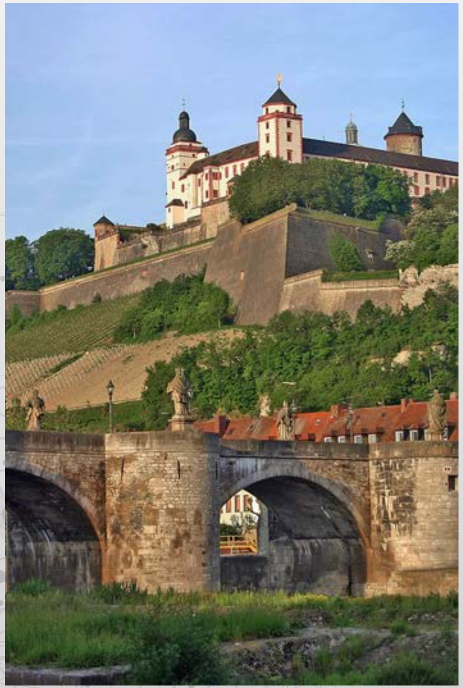
Mario Morvan
Baseline model design

Ingo Waldmann:
Websites and stuff



The whole UCL ExoAI group
For technical & emotional support

Correcting Transiting Exoplanet Light Curves for Stellar Spots



- Ran between April 2019 & 15th August 2019
- An official Discovery Challenge of **ECML-PKDD**
  (Top European **Machine Learning** venue)
- Talk & dedicated special session on **EPSC-DPS**
  (Top European **Planetary Science** venue)
- Goals:
  1) **Explore feasibility** & **identify initial solutions** to a hard problem
  2) **Bridge gap between ML & astrophysics** communities
  3) **Promote ARIEL** to a broader audience (scientific & beyond)
  4) **Establish expertise & infrastructure** to host similar challenges in the future

A natural synergy

Astronomy

- Large, high-dimensional, multimodal datasets
- Low SNR, many sources of noise (systematic / random)
- Complex phenomena to model; parametric form a-priori unknown
- Ill posed problems

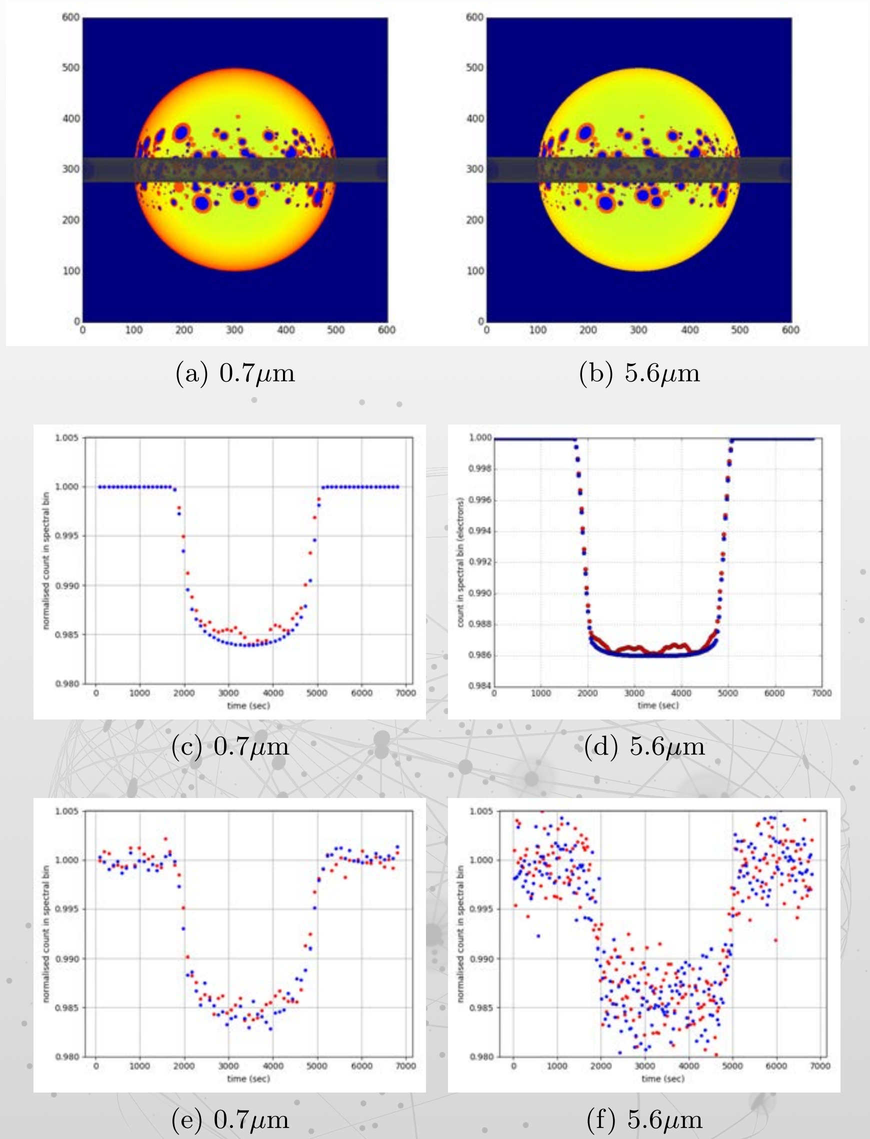
Machine learning

Looking for hard, interesting problems requiring efficient modelling of large and/or complex datasets

Problem: Stellar Spots

To characterise atmospheres need precision (transit depth): 10^{-4} -10^{-5}

Key Question: Is there a way to **efficiently automate transit depth measurements in the presence of stellar noise to the desired precision?**



Sakar et al. 2020


DATASET & BASELINE

The Challenge:
Correct Light Curves for Stellar spots

- 2100 target stars x10 stellar spot x10 photon noise instances = 2.1×10^6 **data points**

- 20% of target stars reserved for testing (**evaluation**)

- **Light curves**, each a **time-series**: relative flux measurements for **300 timesteps** (5h obs.) @ **55 channels** (wavelength bins)

- **6 additional parameters** (all knowable in advance):
  orbital period, stellar temperature, stellar surface gravity,
  stellar radius, stellar mass, stellar K magnitude

- Targets to predict: relative transit depth @ each wavelength (55 real numbers) — **multi-target regression**

The scoring system

To evaluate a new solution (model), the participants are requested to upload its predictions on the test dataset via the Upload page. See the Data Formats section of the Documentation page for information on the upload format. A score will then be automatically calculated for the solution.

The score reported is based on the weighted average of the absolute error per target (i.e. on the relative radii) across all test set examples i and all wavelengths j and is given by:

$$Score = 10^4 - \frac{\sum_{i \in Test} \sum_{j=1}^{55} w_{ij} 2 y_{ij} |\hat{y}_{ij} - y_{ij}|}{\sum_{i \in Test} \sum_{j=1}^{55} w_{ij}} 10^6,$$

where y_{ij} is the true relative radius and $\hat{y}_{ij}$ the predicted relative radius of the j-th wavelength of the i -th test set example and the corresponding weight w_{ij} is given by:

$$w_{ij} = \frac{1}{{\sigma_{ij}}^2 {\delta_{F_{ij}}}^2},$$

with σ_{ij}^2 being the variance of relative stellar flux caused by the observing instrument at the j-th wavelength of the i-th example and $\delta_{F_{ij}}^2$ the variation of the relative stellar flux caused by stellar spots in the j-th wavelength of the i-th example.

The scoring system

Intuition: Each **target** (each observation's rel. transit depth @ each wavelength) is **weighted by its relative difficulty** — due to both photon & stellar spot noise

Only used for evaluation (scoring of solutions)

Weights were unknown to participants

Participants (& baseline solution) used the **unweighted objective** as a **proxy loss**

The Baseline model

- **Preprocessing**:

  Smoothing: moving-median of window size 3

  Clipping: upper bound to 1

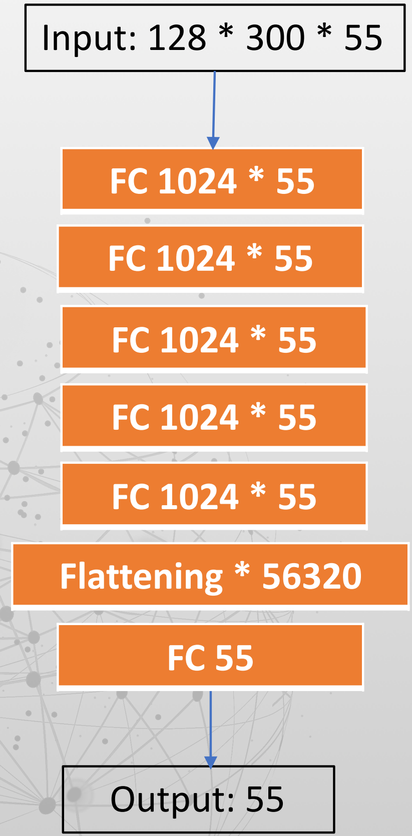
  Normalizing within [-1 , 1]

The Baseline model

- **Fully connected feedforward neural network**
  (Independent prediction of each of the rel. depths of the 55 wavelengths per observation, using only light curves — not the additional parameters provided)

- **Architecture**:
  5 hidden layers (1024 units x 55 channels) followed by a flattening layer

  All ReLU activations except final linear layer of 55 outputs



The Baseline model

- **Training**:

  Initial learning rate 10^{-4} with decay 0.01

  Loss function: Average MSE of all 55 targets
  (i.e. not the one used in evaluation, its unweighed counterpart)

  No batch normalisation, regularization or dropout

  No hyperparameter optimization

  Only 5000 datapoint from the training set used 4020 (train) + 980 (val)

The Baseline model

- Very simple to implement, very fast to train

- Precision within the 10^{-4} region

- A lot of room for improvement:
  explore other models & preprocessing methods
  use more provided datapoints
  use more provided parameters
  share information across wavelengths
  proper hyperparameter tuning
  combining predictions from multiple models
  ….

- Indeed, the participants outperformed our baseline pushing precision to the 10^{-5} region!


OUTCOMES &
PARTICIPATION

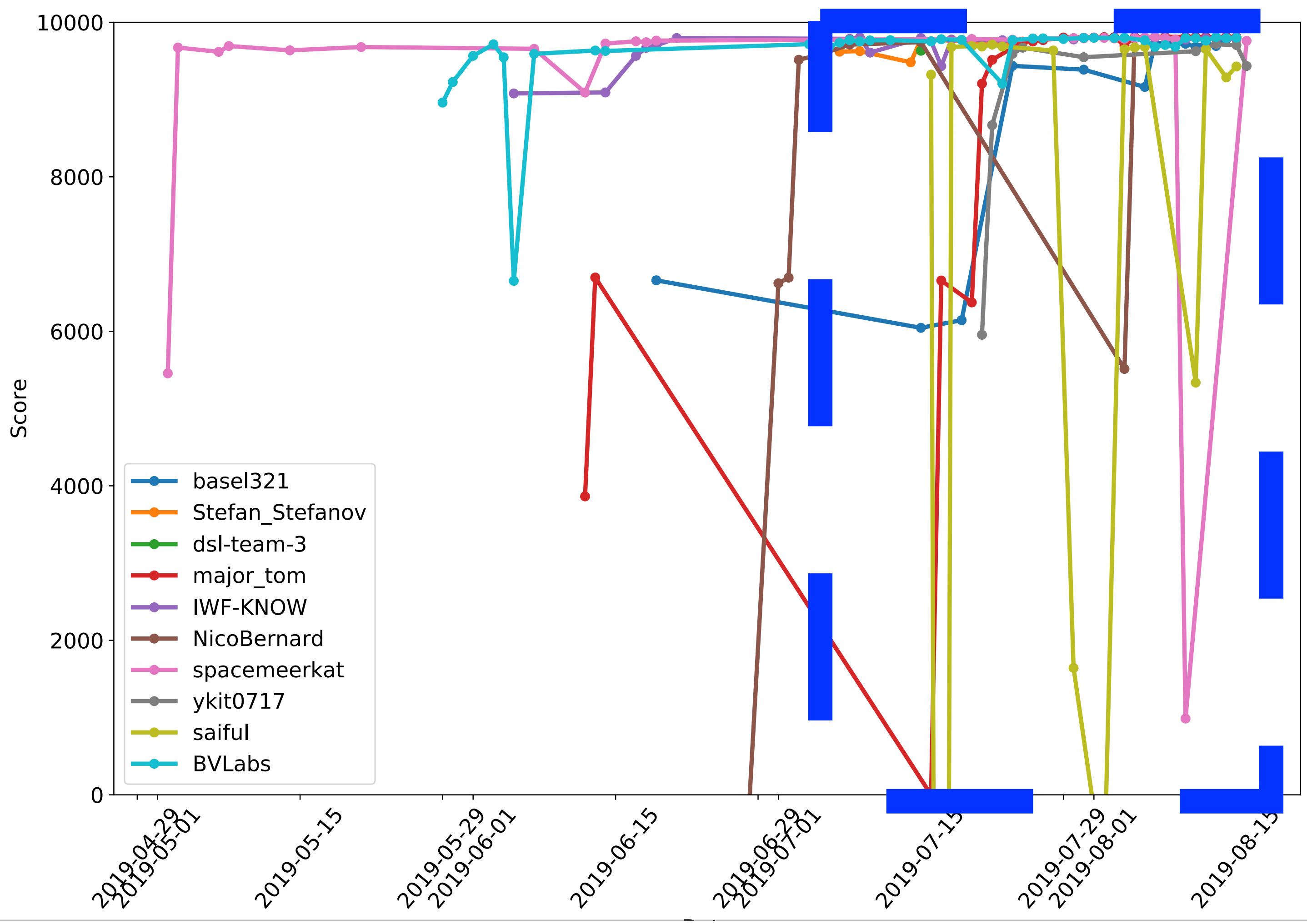
The Final Leaders

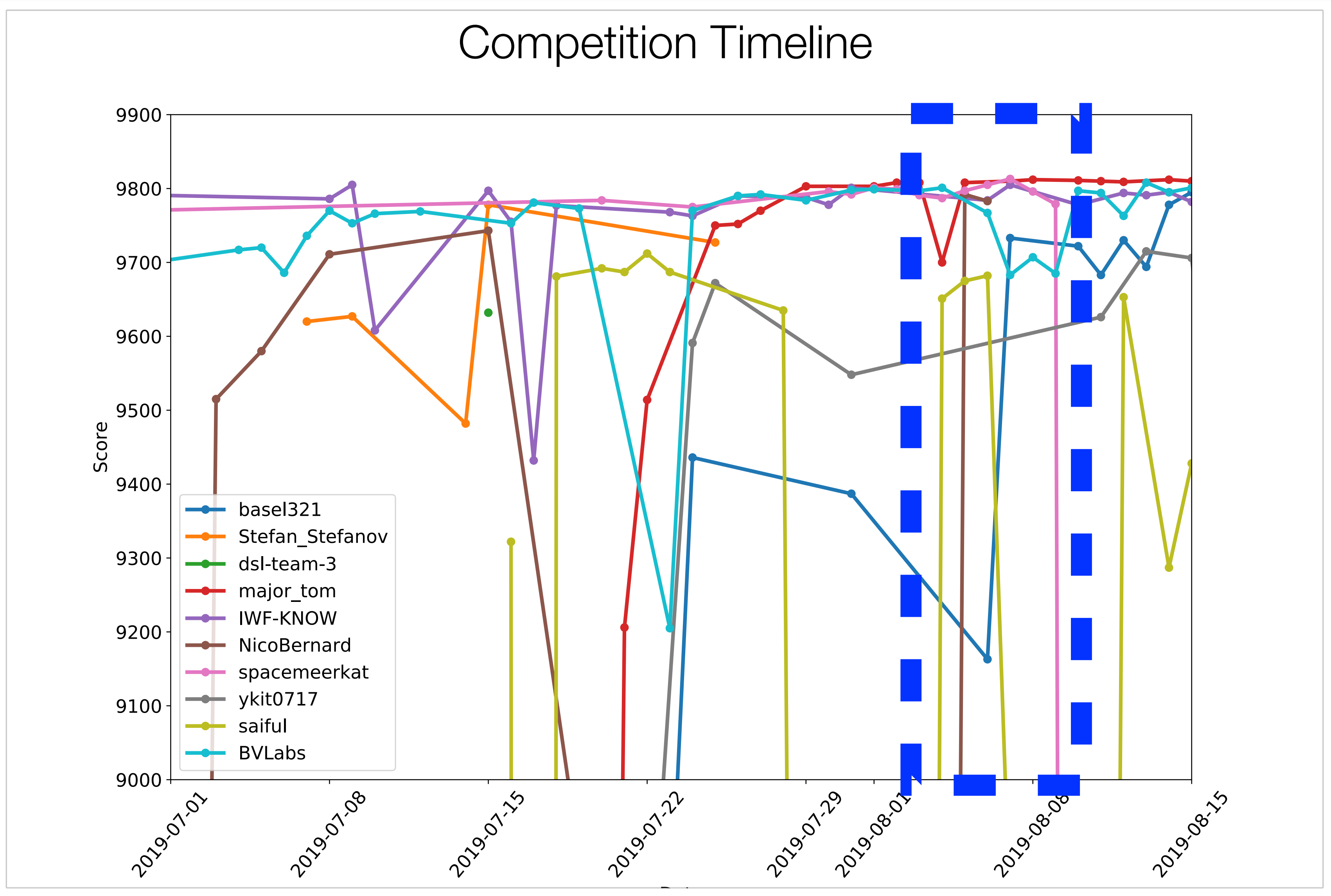
- Overall **112** teams **signed up**, **18** teams **submitted solutions**
  13 teams beat the baseline!
- It was **extremely close** & **differences of <10 are likely mostly due to statistical noise / slightly overfitting to test set**
- First 5 leaders invited to present in ECML & to a joint publication (in progress)
- First 2 leaders won an ECML registration / €590 equivalent
- 1st: **SpaceMeerkat**
  a.k.a. **James Dawson**
- Close 2nd: **major_tom**
  a.k.a. **Vadim Borisov**

Current Leader Board

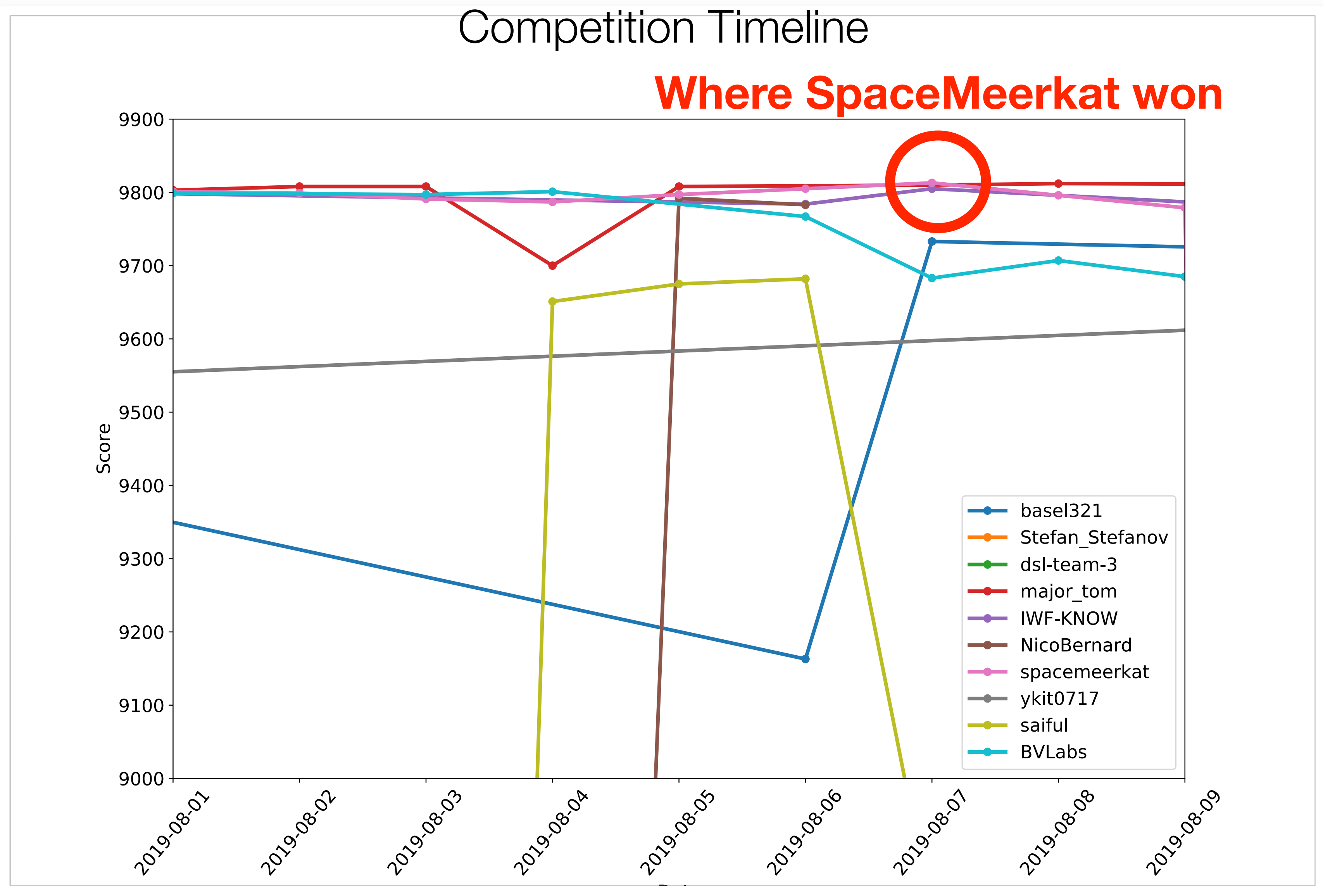
Rank	Name	Score
1	SpaceMeerkat	9813
2	major_tom	9812
3	BVLabs	9808
4	IWF-KNOW	9805
5	basel321	9795
6	NicoBernard	9792
7	Stefan_Stefanov	9778
8	ykit0717	9715
9	saiful	9712
10	dsl-team-3	9632

Competition Timeline





Competition Timeline
9900
9800
9700
9600
9500
9400
9300
9200
9100
9000
Score
basel321
Stefan_Stefanov
dsl-team-3
major_tom
IWF-KNOW
NicoBernard
spacemeerkat
ykit0717
saiful
BVLabs
2019-07-01
2019-07-08
2019-07-15
2019-07-22
2019-07-29
2019-08-01
2019-08-08
2019-08-15



Competition Timeline
Where SpaceMeerkat won
9900
9800
9700
9600
9500
9400
9300
9200
9100
9000
Score
2019-08-01
2019-08-02
2019-08-03
2019-08-04
2019-08-05
2019-08-06
2019-08-07
2019-08-08
2019-08-09
basel321
Stefan_Stefanov
dsl-team-3
major_tom
IWF-KNOW
NicoBernard
spacemeerkat
ykit0717
saiful
BVLabs

The Final Leaders

- We will have the chance to hear both **SpaceMeerkat** a.k.a. **James Dawson** **major_tom** a.k.a. **Vadim Borisov** presenting their solutions in this session
- We will also have the chance of hearing another participant, **Artash Nath** presenting his…
- Artash is **an 8th Grade student (!)** from Toronto; his participation is a testament both to his skill and love for learning and to the success of the competition as a means of promoting ARIEL to a broader, international audience & a tool for knowledge exchange and interdisciplinary problem solving!

Current Leader Board

Rank	Name	Score
1	SpaceMeerkat	9813
2	major_tom	9812
3	BVLabs	9808
4	IWF-KNOW	9805
5	basel321	9795
6	NicoBernard	9792
7	Stefan_Stefanov	9778
8	ykit0717	9715
9	saiful	9712
10	dsl-team-3	9632

Interesting observations

- We will now briefly present some interesting observations drawn from the top 5 winning solutions:

Current Leader Board

Rank	Name	Score
1	SpaceMeerkat	9813
2	major_tom	9812
3	BVLabs	9808
4	IWF-KNOW	9805
5	basel321	9795

Interesting observations

1) All top-5 participants **used methods capable of producing highly non-linear models**: **random forests**, **gradient boosting** (**BV-labs**) and **deep neural networks** (**all top-5 teams**)

Unless the feature space is appropriately transformed (this requires domain knowledge), indeed the original inputs (raw light curves & additional parameters) need to be nonlinearly combined to meaningfully predict the rel. transit depth in the presence of stellar spots

Interesting observations

2) That said, **by clever (i.e. informed by physics) transformation of the feature space (feature engineering)**, it is **possible to come up with competitive models that are linear** in the transformed feature space:

e.g. team **IWF-KNOW** produced several models, several of which are **linear in parameters inspired by**:

Seager, Sara, and Gabriela Mallen-Ornelas. ”A unique solution of planet and star parameters from an extrasolar planet transit light curve.” The Astrophysical Journal 585.2 (2003): 1038.APA

Interesting observations

3) Most participants did not just train non-linear models, but went a step further: they **trained multiple and combined their predictions** (i.e. formed ‘**ensembles**’ of predictors)

Methods capable of producing highly non-linear models are often prone to overfitting the data; by training multiple models and aggregating their predictions, the effect of overfitting (what is called in statistics the ‘variance component of the expected error’) is reduced.

Interesting observations

4) The top-ranked participant (**James Dawson**) noticed that **the training dataset contained a few outliers**; by removing them, overall performance of the models he trained increased

The outliers were not due to dataset generation issues (but rather due to 1 target having an uncharacteristically large rel. transit depth). This only appeared in the training and not the test data, therefore methods trained to fit this target would not generalise well to typical cases.

Interesting observations

5) All top-ranked participants gave **considerably more emphasis to the light curve data than the additional parameters**; Most report that **their addition only marginally increased their scores**

This might suggest that the additional parameters share a lot of common information with the light curves themselves and might be redundant.

<u>Additional parameters:</u>

orbital period,
stellar temperature,
stellar surface gravity,
stellar radius,
stellar mass,
stellar K magnitude

Before our winner’s presentations,

some closing remarks….

Closing remarks: were the goals met?

1) **Explore feasibility** & **identify initial solutions** to a hard problem

Baseline & most entries **achieved desired precision — efficient automation of solution is feasible**

Most successful approaches used **ensembles of DNNs / CNNs** (e.g. **all top-5 winners**, **Artash Nath**)

But **simpler approaches competitive**
e.g. team **IWF-KNOW proposed** —beyond their other solutions—
a **linear model in an appropriately transformed parameter space**

Closing remarks: were the goals met?

2) **Bridge gap between ML & astrophysics** communities

Strong presence in top European venues of both fields (ECML-PKDD & EPSC-DPS)

The 113 participants spanned both fields

Solutions included both **astrophysicists using ML methods** (e.g. **James Dawson**)

…and **ML researchers using astrophysics domain knowledge** (e.g. **BV-labs**)

Top-5 teams are collaborating with us on a joint publication (stay tuned)

Closing remarks: were the goals met?

3) **Promote ARIEL** to a broader audience (scientific & beyond)

Participation large (123 teams) — rivaled Kaggle Challenges & **diverse**

Sectors: academia, industry, amateurs

Scientific fields: astrophysics, machine learning, statistics

Location: not limited to Europe, but international

Artash is a prime example of the competition's reach
and potential to attract more interest and talent to ARIEL!

Closing remarks: were the goals met?

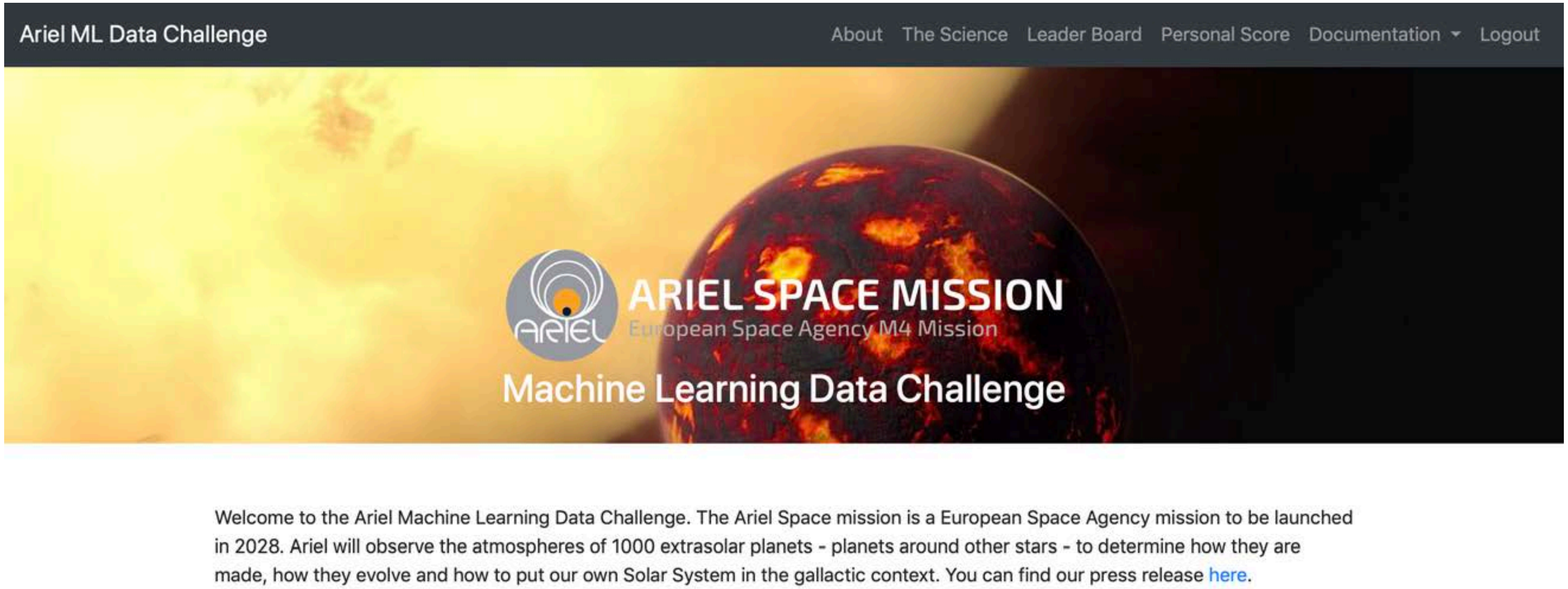
4) **Establish expertise & infrastructure** to host similar challenges in the future.

Dedicated **website** developed from scratch

A **ruleset** for the competition was established — a roadmap for future competitions

Lessons learned: e.g. next time **leader board score** will be calculated on **only part of the test set** to limit overfitting to test set even further

Our team was **trained** in organising machine learning challenges, advertising them in the appropriate venues using the proper terminology, coordinating the different aspects, foreseeing potential issues, …



The data challenge website remains accessible at:

http://ariel-datachallenge.space

It has been repurposed as a data archive service and the full training set will be archived with a permanent DOI


PRESENTATIONS
BY OUR WINNERS
ARIEL
europlanet
SOCIETY
UK SPACE
AGENCY
Horizon 2020
erc
European Research Council
Established by the European Commission
UCL
CENTRE FOR SPACE
EXOPLANET DATA