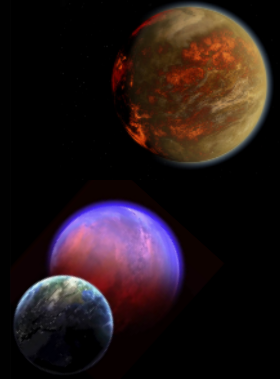
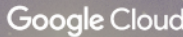
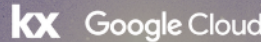


Machine Learning for Exoplanet research



Lessons and results from
NASA Frontier Development Lab

Dr. Daniel Angerhausen
CSH Fellow
Bern University





Sagan School 2009

Who am I?

u^b
UNIVERSITÄT
BERN
CSH
CENTER FOR SPACE AND
HABITABILITY

explainables
SCIENCE COMMUNICATION

PlanetS

10 Blue Marble Space
Institute of Science
Celebrating 10 years of exploration

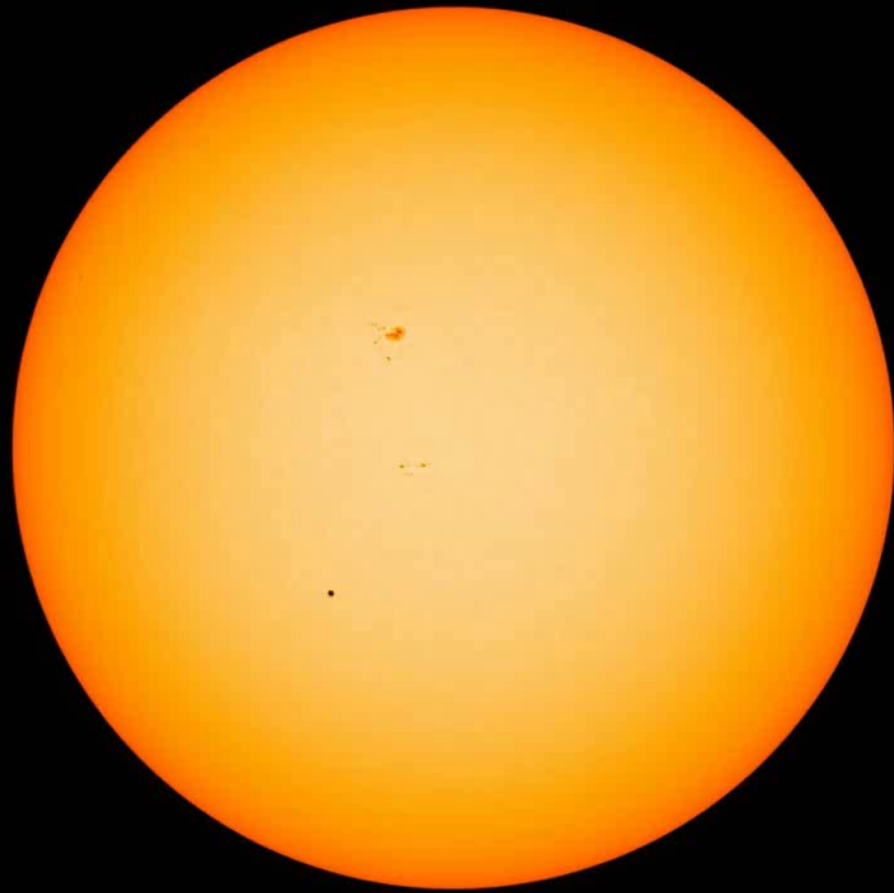
?ETI
INSTITUTE

Google Cloud

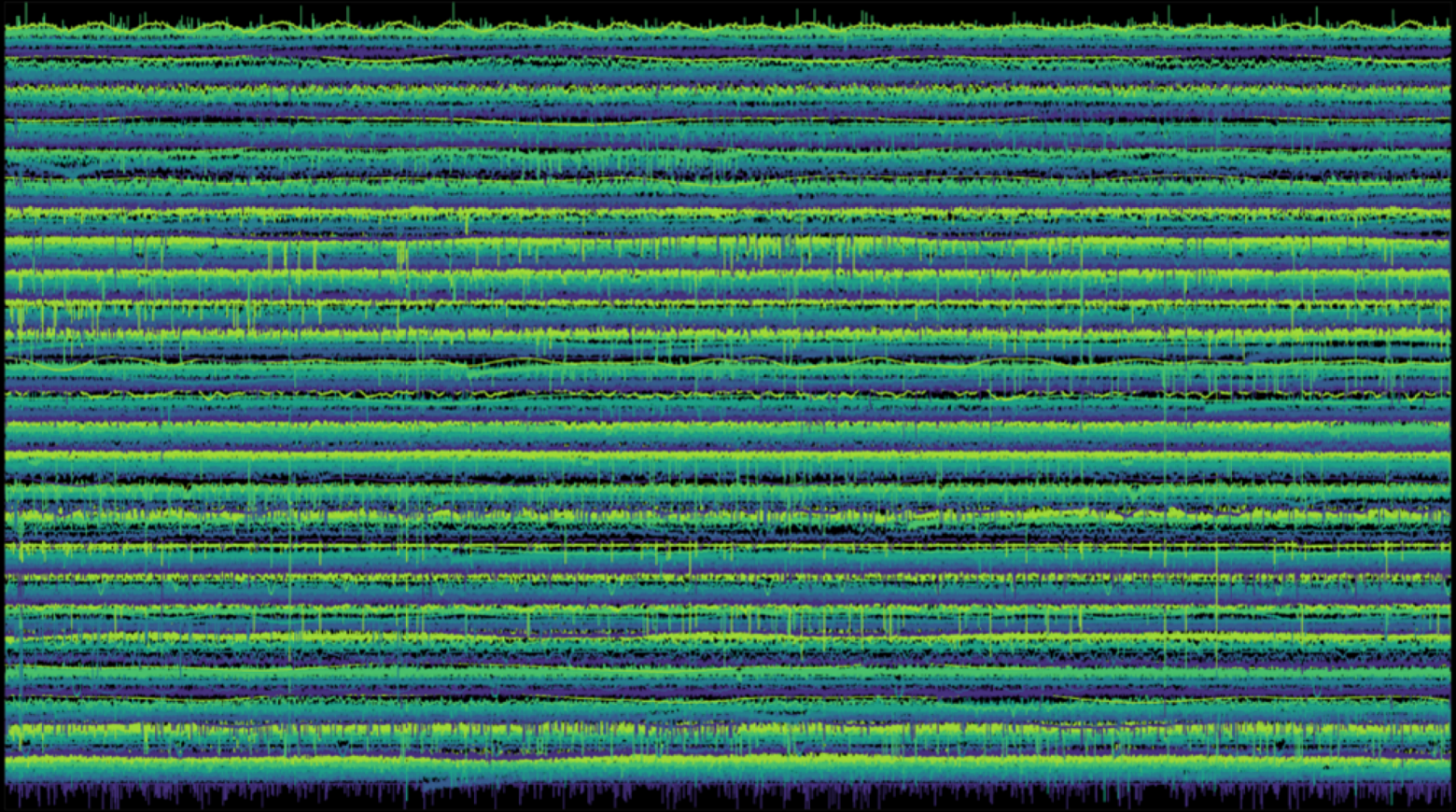
NASA

NASA
FRONTIER
DEVELOPMENT LAB





A mercury sized planet transiting a sun-like star



Data > Information > Knowledge > Wisdom

Artificial Intelligence

```
graph TD; AI[Artificial Intelligence] --- ML[Machine Learning]; ML --- CA[Classic algorithms]; ML --- NN[Neural Networks]; NN --- DL[Deep Learning]
```

Machine Learning

Classic
algorithms

Neural Networks

Deep
Learning



playground.tensorflow.org

Tinker With a **Neural Network** Right Here in Your Browser.
Don't Worry, You Can't Break It. We Promise.



Epoch
000,438

Learning rate
0.03

Activation
Tanh

Regularization
None

Regularization rate
0.3

Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%



Noise: 5



Batch size: 10



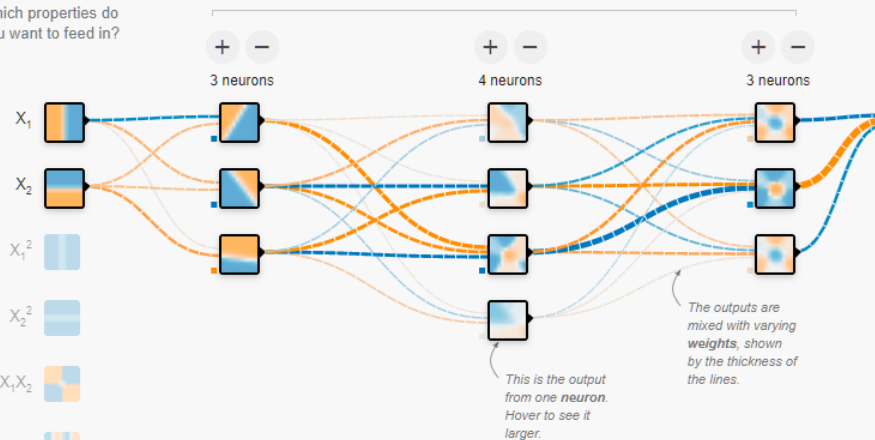
REGENERATE

FEATURES

Which properties do you want to feed in?

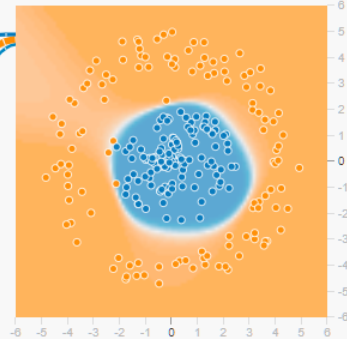
- X_1
- X_2
- X_1^2
- X_2^2
- $X_1 X_2$
- $\sin(X_1)$
- $\sin(X_2)$

+ - 3 HIDDEN LAYERS



OUTPUT

Test loss 0.029
Training loss 0.012



Colors shows data, neuron and weight values.

“FDL is an **applied AI research accelerator** established to maximize new AI technologies and capacities emerging in **academia and the private sector** and apply them to **challenges in the space sciences.**”



The 2018 Challenges

Defined by NASA and carefully curated group of space scientists, humanitarians and technologists in a “Big Think”

#AIforGOOD



The 2019 Challenges



The Recipe

Example: FDL 2018 Astrobiology Team



Michael Himes
Planetary Scientist



Molly O'Beirne
Planetary Scientist



Shawn Domagal-Goldman
Planetary Mentor



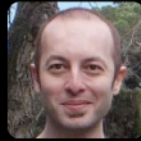
Giada Arney
Planetary Mentor



Frank Soboczinski
Computer Scientist



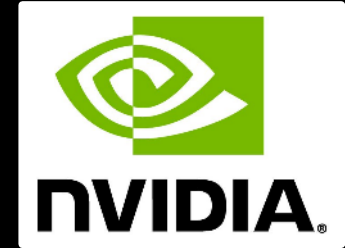
Simone Zorzan
Computer Scientist



Atilim Gunes Baydin
AI Mentor



Daniel Angerhausen
Planetary Mentor

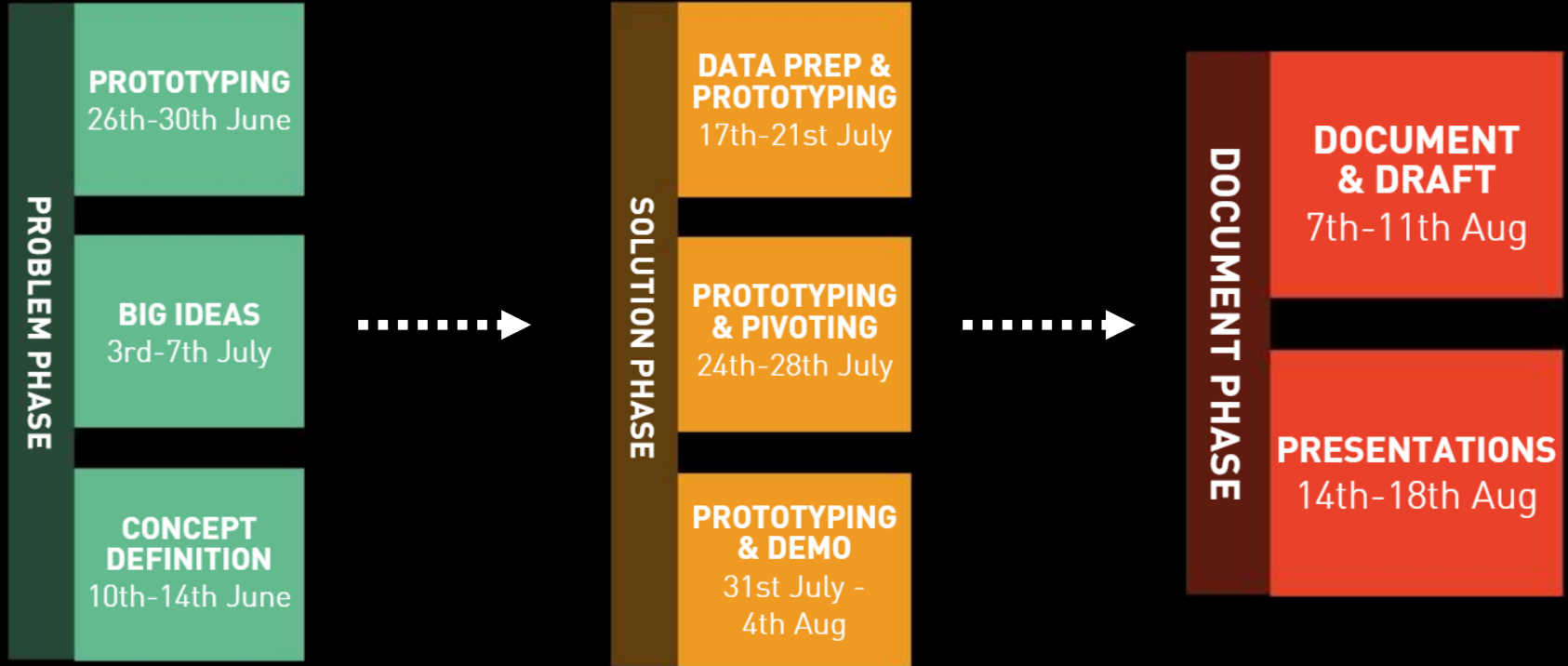


Participants + Mentors + Silicon Valley

**And then lock them up
at the SETI Institute for
8 weeks...**



...the Silicon Valley way





Can we use AI techniques for localization on the Moon?

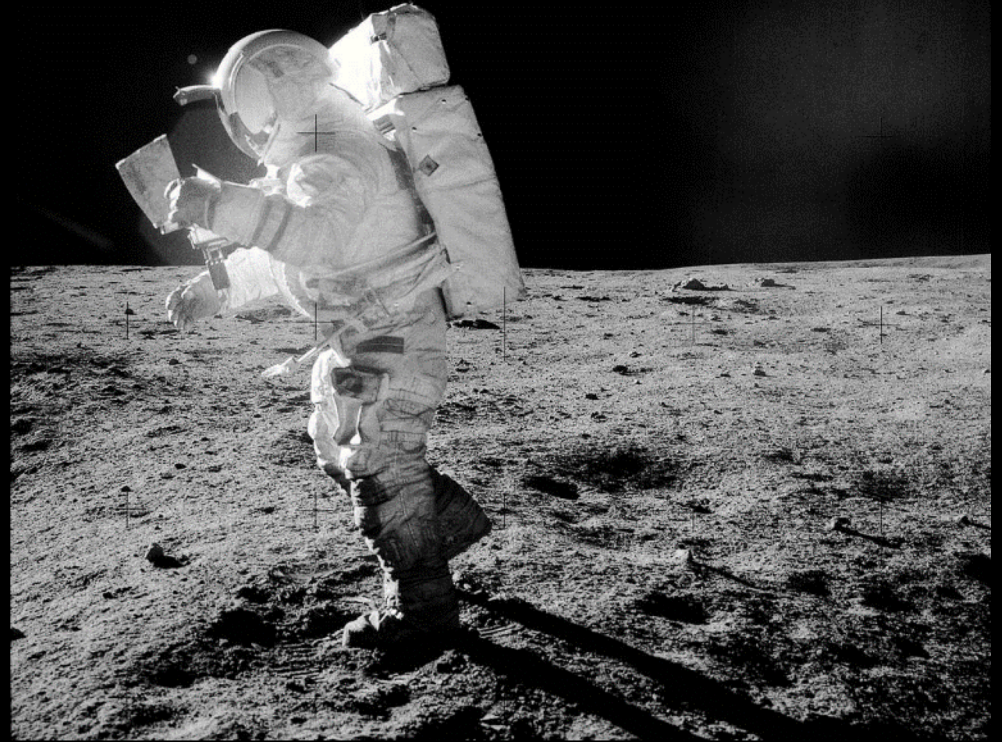
Andrew Chung, Philippe Ludivig, Ben Wu,
Ross Potter



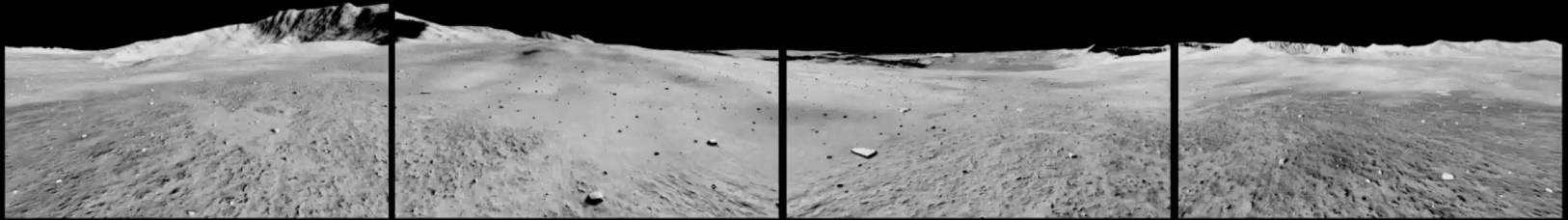
The Problem: Where are you?



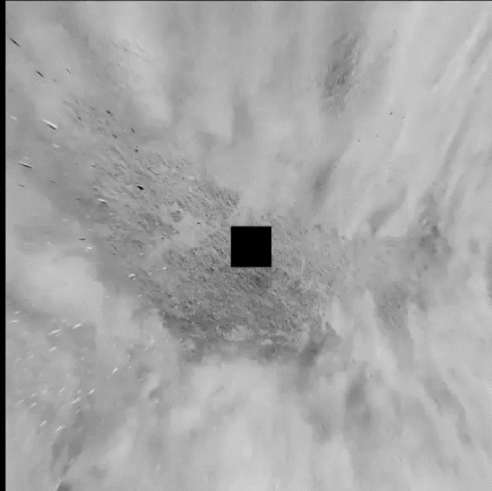
The Problem: Where are you?



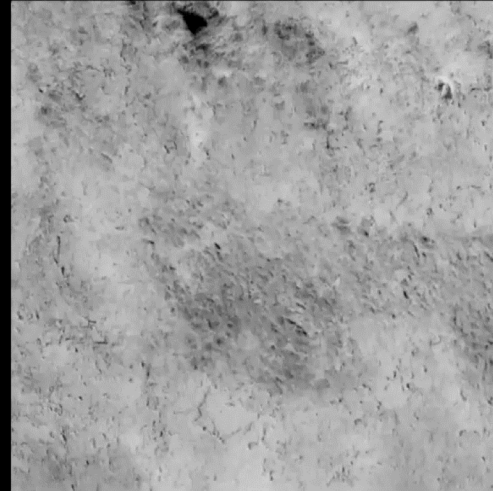
The Problem



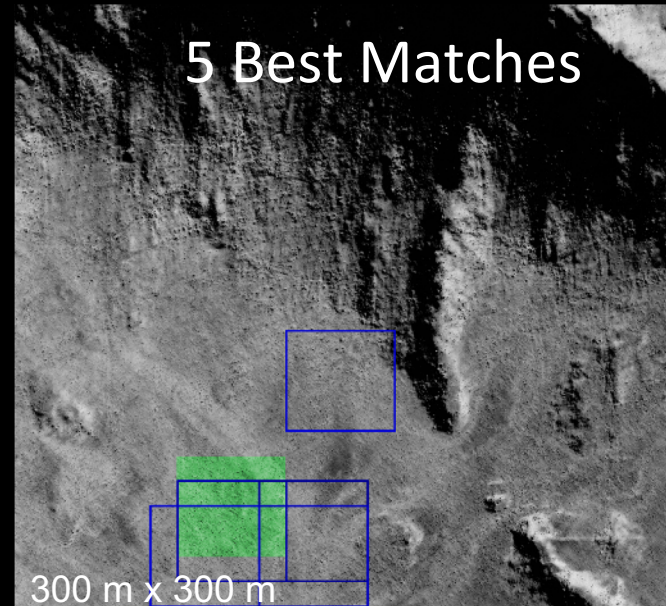
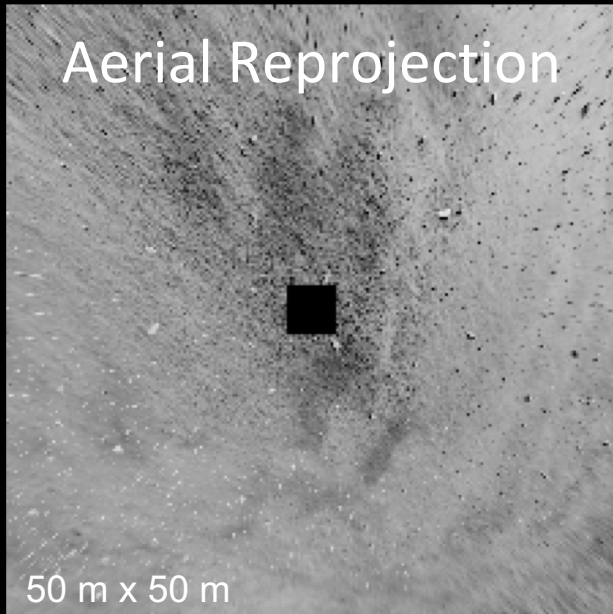
Reprojected
View



True Orbital
View



The Breakthrough



Increase the efficacy and yield of **exoplanet** **transit detections** with **deep learning**

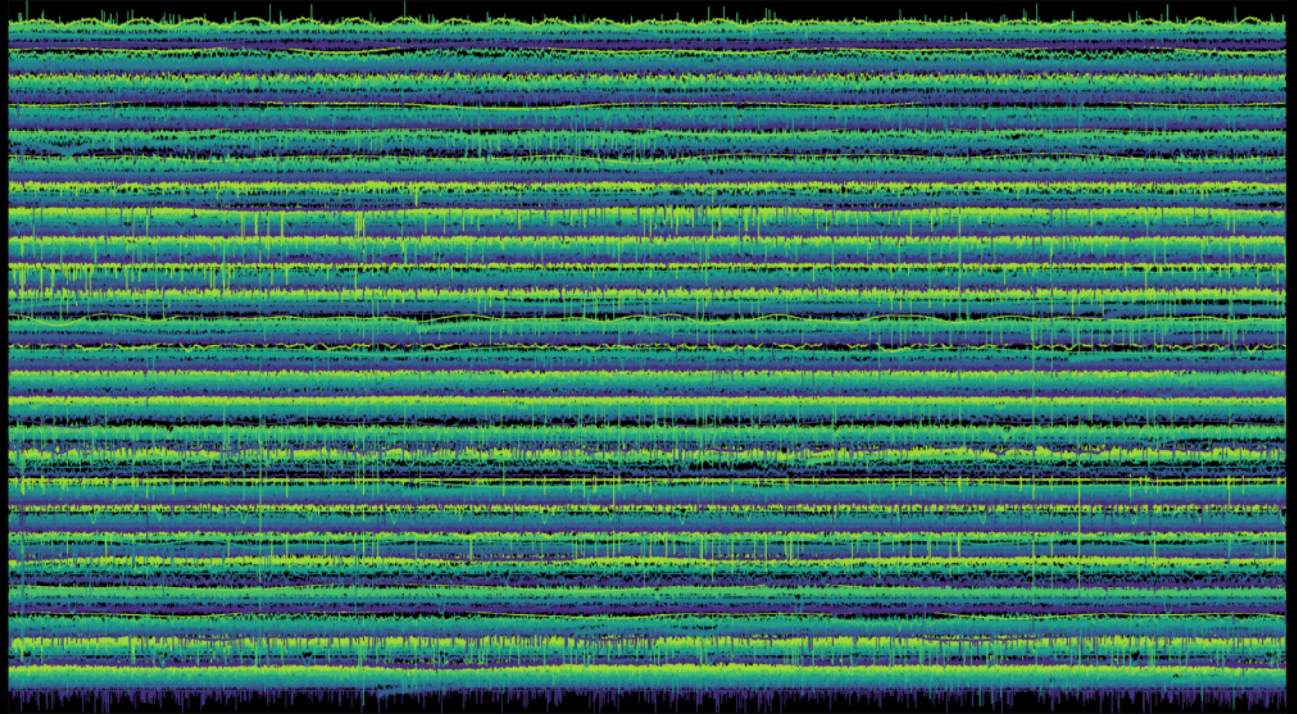
Michele Sasdelli, Megan Ansdell, Hugh
Osborn, Yani Ionnou



Google Cloud **kx**

The Problem

Where are
the planets
and are they
real?



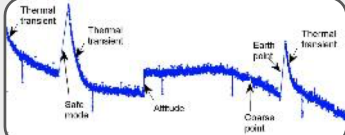
Kepler/TESS Pipelines

Target Pixel File (TPF)



Smith+2012, Stumpe+2012

Aperture Photometry & Systematics Correction



Jenkins+2010, Seader+2013

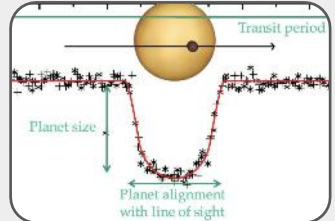
Transiting Planet Search (TPS)



Threshold Crossing Event (TCE)

Wu+2010

Data Validation (DV)



Exoplanet Catalogues

Star Name	Planet Name	Discovery Method	Planet Size (Earth radii)	Orbital Period (days)	Distance (light years)
Kepler-90	Kepler-90b	Transit	1.03	13.1017	2040
Kepler-90	Kepler-90c	Transit	1.03	305.987	2040
Kepler-90	Kepler-90d	Transit	1.03	385.272	2040
Kepler-90	Kepler-90e	Transit	1.03	451.305	2040
Kepler-90	Kepler-90f	Transit	1.03	510.976	2040
Kepler-90	Kepler-90g	Transit	1.03	560.595	2040
Kepler-90	Kepler-90h	Transit	1.03	610.164	2040
Kepler-90	Kepler-90i	Transit	1.03	659.733	2040
Kepler-90	Kepler-90j	Transit	1.03	709.302	2040
Kepler-90	Kepler-90k	Transit	1.03	758.871	2040
Kepler-90	Kepler-90l	Transit	1.03	808.440	2040
Kepler-90	Kepler-90m	Transit	1.03	858.009	2040
Kepler-90	Kepler-90n	Transit	1.03	907.578	2040
Kepler-90	Kepler-90o	Transit	1.03	957.147	2040
Kepler-90	Kepler-90p	Transit	1.03	1006.716	2040
Kepler-90	Kepler-90q	Transit	1.03	1056.285	2040
Kepler-90	Kepler-90r	Transit	1.03	1105.854	2040
Kepler-90	Kepler-90s	Transit	1.03	1155.423	2040
Kepler-90	Kepler-90t	Transit	1.03	1204.992	2040
Kepler-90	Kepler-90u	Transit	1.03	1254.561	2040
Kepler-90	Kepler-90v	Transit	1.03	1304.130	2040
Kepler-90	Kepler-90w	Transit	1.03	1353.699	2040
Kepler-90	Kepler-90x	Transit	1.03	1403.268	2040
Kepler-90	Kepler-90y	Transit	1.03	1452.837	2040
Kepler-90	Kepler-90z	Transit	1.03	1502.406	2040

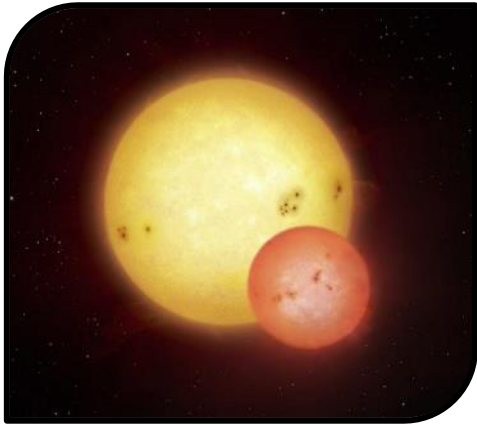
Batalha+2013, Burke+2014, Rowe+2015, Mullally+2015

Kepler TCE Review Team [human vetting]

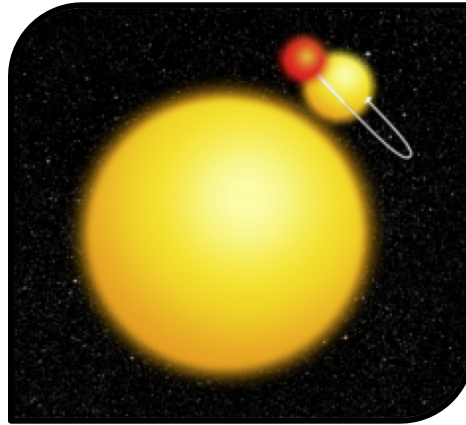


Where deep learning can (and is!) helping

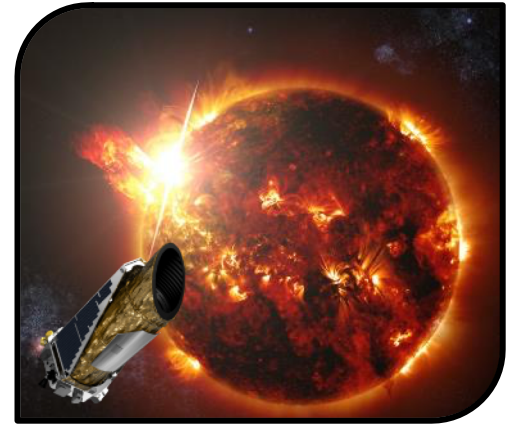
The Data: False Positives



Eclipsing Binaries (EBs)

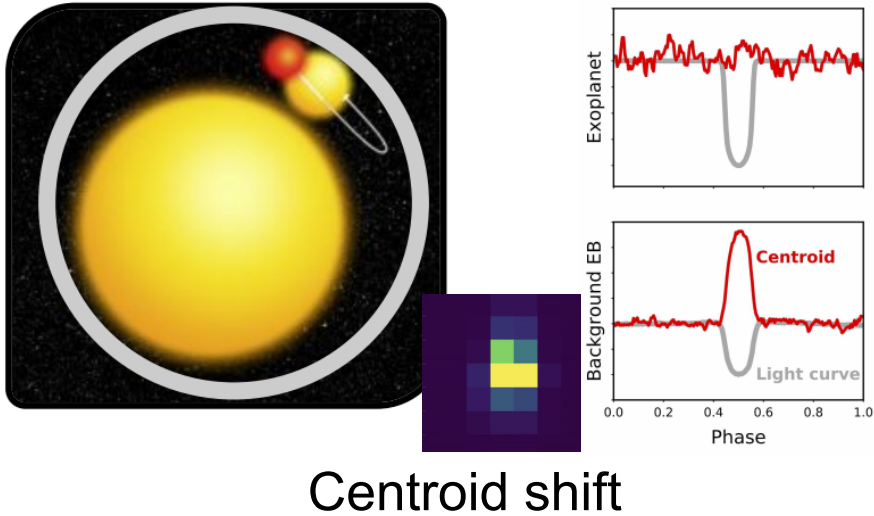


Background Eclipsing Binaries (BEBs)

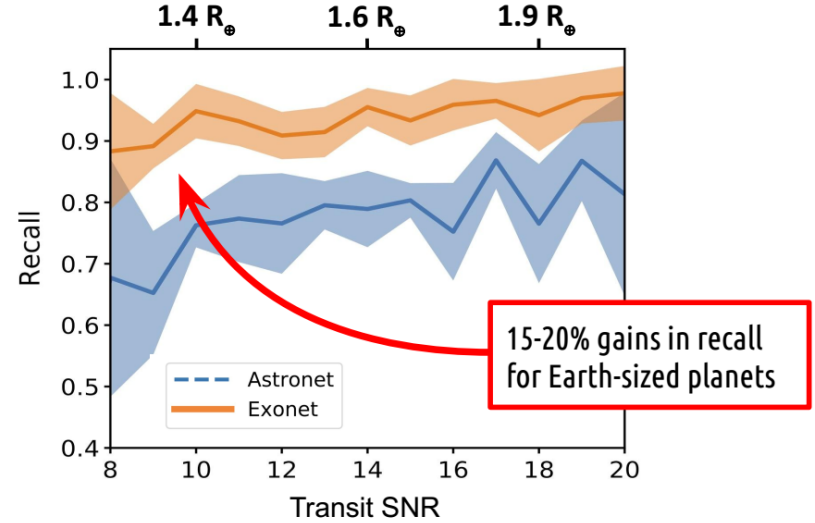


Stellar Variability / Instrumental Noise

Add domain knowledge



Improved model



Details: Ansdell et al. 2018, Osborn et al. 2019

Challenge 1: Understanding what is universally possible for life

Aditya Chopra, Aaron Bell, William Fawcett, Rodd Talebi

Challenge 2: From biohints to confirmed evidence of life

Michael Himes, Frank Soboczinski, Simone Zorzan, Molly O'Beirne



The Problem

N=1

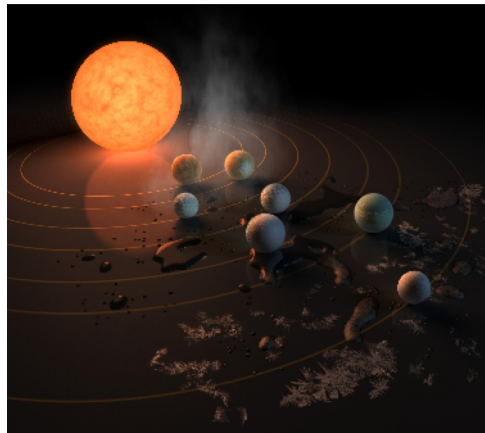
...not exactly BigData



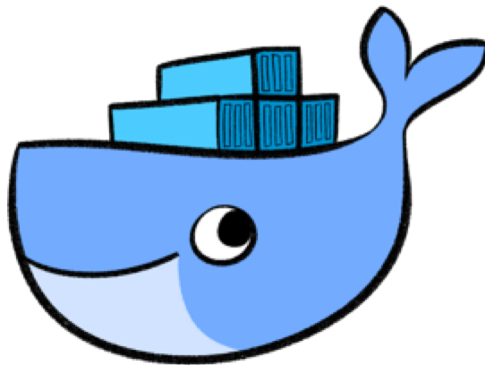
Team 1
~200.000
ATMOSpheres
Calculated



vpl Atmos



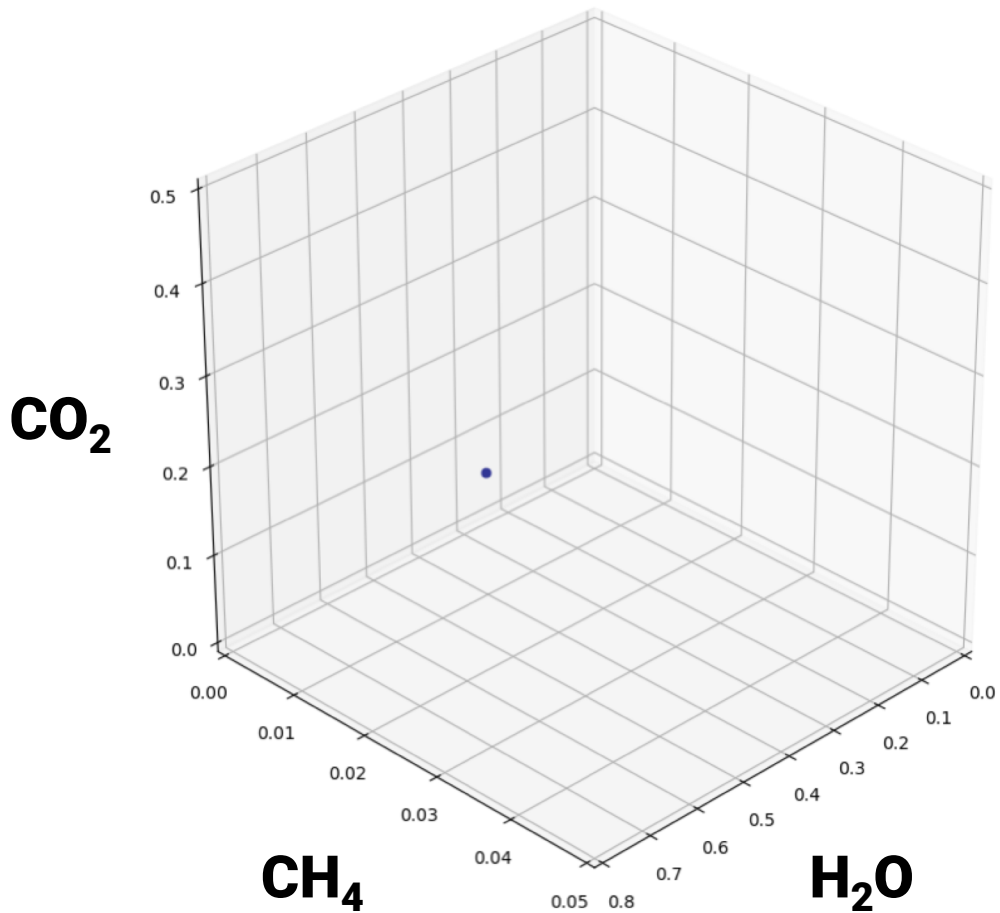
docker



pyATMOS

 **Google Cloud**





Team 2

3 Million

**Observations
of spectra
simulated**



Team 2

3 Million

**Observations
of spectra
simulated**



Google Cloud



The breakthrough

Datasets and software soon available:

NASA Exoplanet Archive

Google cloud/Kaggle

NASA EXOPLANET ARCHIVE
NASA EXOPLANET SCIENCE INSTITUTE

Home About Us Data Tools Support Login

Frontier Development Labs (FDL) PyATMOS Dataset

Summary of Atmospheric Models

Download All Checked Models

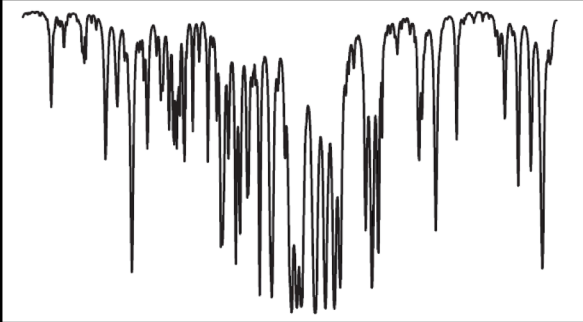
	Input CH ₄ Concentration (Fraction)	Input CO ₂ Concentration (Fraction)	Input H ₂ Concentration (Fraction)	Input H ₂ O Concentration (Fraction)	Input O ₂ Concentration (Fraction)	Pressure (bar)	Temperature (K)
EP	0.01000000	0.03000000	0.05000000	0.84000000	0.22000000	1.06290000	320.29000000
HE	0.02000000	0.01000000	0.00000000	0.81000000	0.16000000	0.81200000	320.26000000
HE	0.02000000	0.01000000	0.00000000	0.80000000	0.21000000	1.11030000	320.21000000
HA	0.00001163	0.04000000	0.00000000	0.15000000	0.14000000	1.06050000	312.87000000
HE	0.02000000	0.02000000	0.00000000	0.80000000	0.20000000	1.11770000	312.48000000
KE	0.01500000	0.01000000	0.00000000	0.81000000	0.35000000	1.09800000	318.51000000
HE	0.00001163	0.00000000	0.00000000	0.85000000	0.14000000	1.02940000	298.59000000
KE	0.03000000	0.02000000	0.05000000	0.80000000	0.25000000	1.13550000	322.82000000

Preview of Selected Model

Download This Model

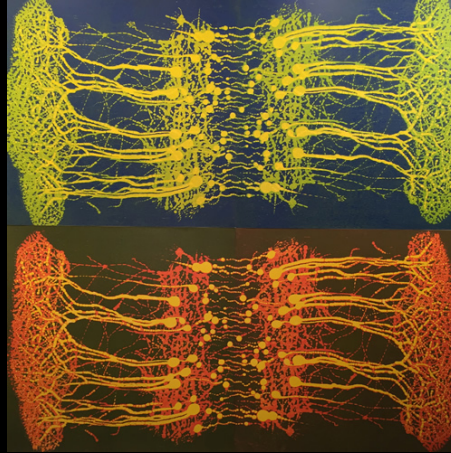
Layer Number	Pressure (bar)	Altitude (km)	Temperature (K)	Is Layer Convective?	H ₂ O Fraction	O ₂ Fraction
1	7.49990E-06	6.72470E+01	8.98150E+01	0	1.16340E-02	6.42840E-09
2	9.31000E-06	6.67220E+01	1.01380E+02	0	4.65610E-03	6.37940E-09
3	1.12410E-05	6.61480E+01	1.09380E+02	0	4.00000E-06	6.31500E-09
4	1.35350E-05	6.55490E+01	1.14480E+02	0	4.00000E-06	6.25970E-09
5	1.61310E-05	6.49230E+01	1.17940E+02	0	4.00000E-06	6.18800E-09
6	1.96130E-05	6.42920E+01	1.19700E+02	0	4.00000E-06	6.07480E-09
7	2.35440E-05	6.36390E+01	1.20580E+02	0	4.00000E-06	5.99600E-09
8	2.82230E-05	6.30280E+01	1.21010E+02	0	4.00000E-06	5.93940E-09

Spectrum



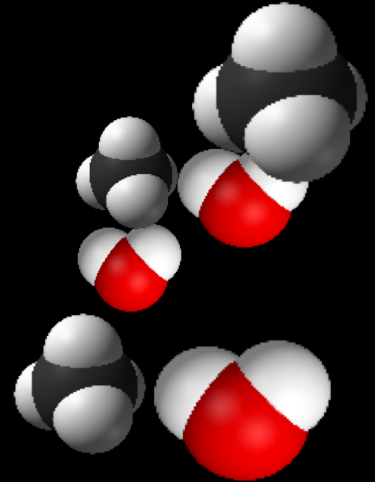
>

DL



>

Composition



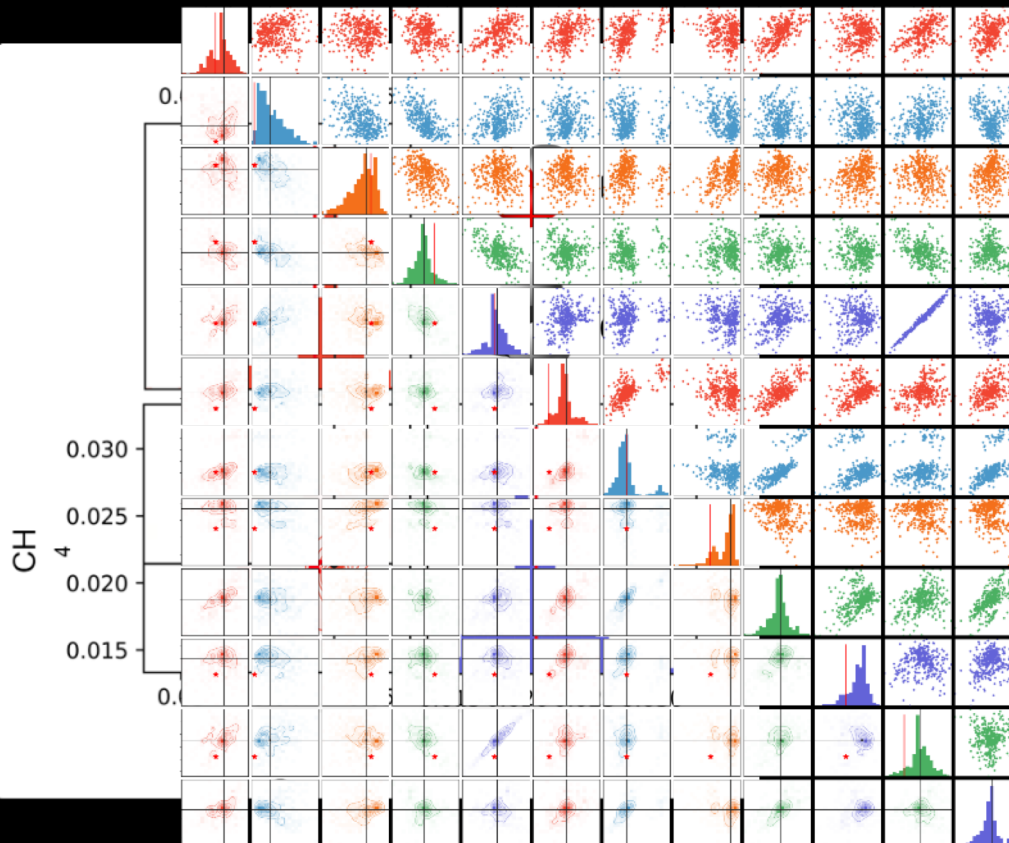
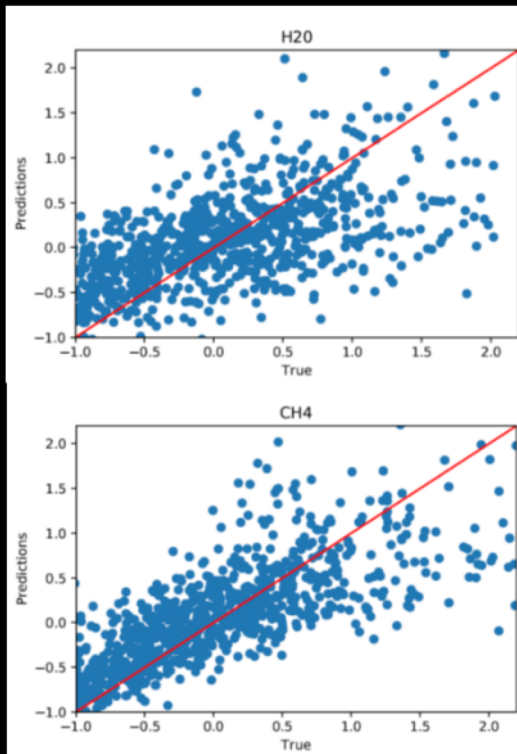
First application of (Ensembles of Bayesian)

Neural Networks in Exoplanet Spectral Retrieval

Soboczenski+ 2018 NIPS, Cobb+ 2019 AJ, Himes+ 2019 in prep

**Can we replace “traditional”
bayesian sampling with neural
networks?**

Yes, but...



Comparison

INARA

Method	Time	# of Molecules Retrieved
Traditional	Hours to days	User-specified
ExoGAN	Minutes	4 (Hot Jupiters)
HELA	Seconds	3 (1 specific Hot Jupiter)
INARA	Seconds	12 (rocky planets)

Comparison

WFC3 spectrum of WASP-12b

Cobb+ 2019 (plan-net, BNN),
Marquez-Neila+ 2018 (HELA, RF)

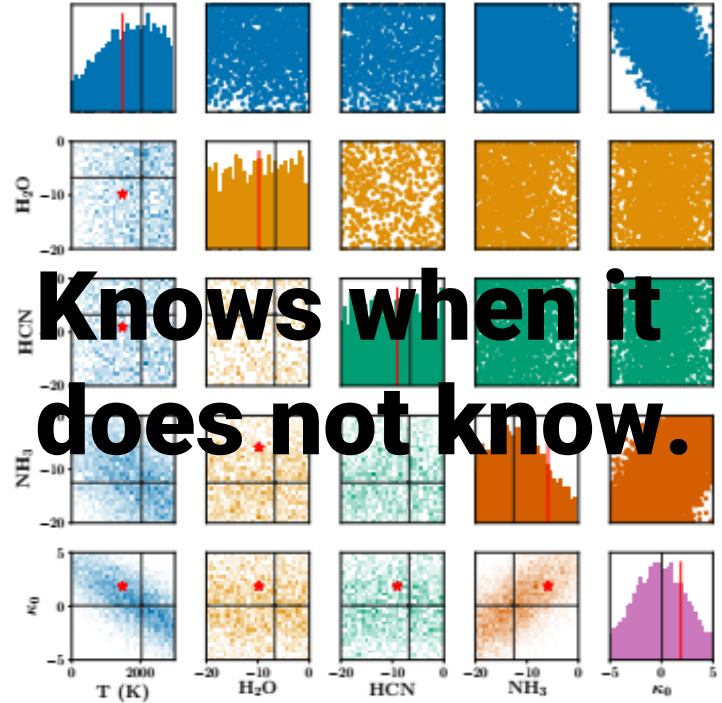
	$T(K)$	$\log X_{H_2O}$	$\log X_{HCN}$	$\log X_{NH_3}$	κ_0	MEAN
PLAN-NET R^2	0.770	0.623	0.487	0.721	0.750	0.673
ENS. 5 PLAN-NET R^2	0.770	0.629	0.491	0.723	0.751	0.673
OUR RAN. FOREST R^2	0.746	0.608	0.466	0.700	0.736	0.651
RAN. FOREST ^a R^2	0.746	0.608	0.467	0.700	0.737	0.652

	$T(K)$	$\log X_{H_2O}$	$\log X_{HCN}$	$\log X_{NH_3}$	κ_0
KREIDBERG ET AL. (2015)	1371 ⁺⁴⁶⁶ ₋₃₄₃	-2.7 ^{+1.0} _{-1.1}	-	-	-
MÁRQUEZ-NEILA ET AL. (2018) NESTED SAMPLING	1105 ⁺⁵⁴⁵ ₋₂₈₇	-3.0 ^{+2.0} _{-1.9}	-8.5 ^{+3.8} _{-2.9}	-8.4 ^{+3.1} _{-2.9}	-2.8 ± 0.9
OUR RAND. FOREST	937 ⁺⁴¹⁰ ₋₁₄₆	-2.835 ^{+1.51} _{-3.37}	-7.484 ^{+3.43} _{-2.89}	-9.202 ^{+4.12} _{-2.74}	-2.281 ^{+1.09} _{-1.57}
ENS. 5 PLAN-NET	1142 ± 412	-2.781 ± 0.429	-8.210 ± 12.7	-9.605 ± 6.7	-2.601 ± 1.23



**Very confident,
but wrong.**

(a) Random Forest

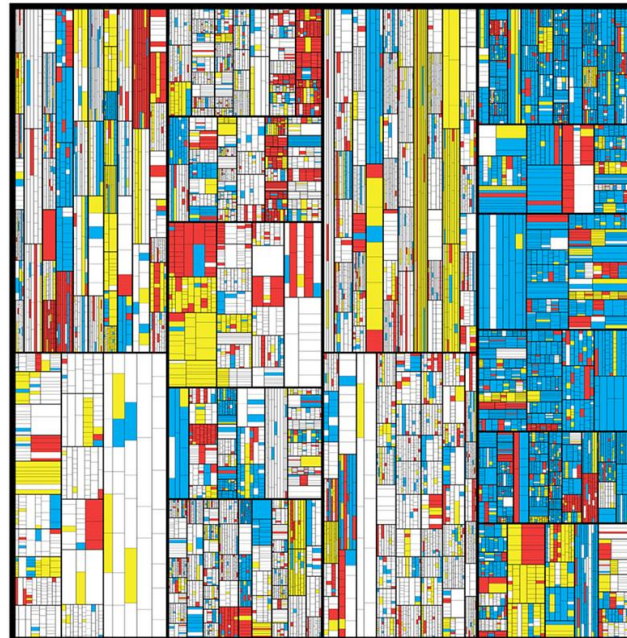


**Knows when it
does not know.**

(b) plan-net Ensemble

Conclusion

- You are already doing it
- Data is the driver
- AI/ML a toolbox, not one hammer
- DL new tool with some applications
- ask an Expert, collaborate



3,028 digits of π

science + art - martin krzywinski

When If we find the *first signs of life in space*:

... machine learning was used

... within a public private partnership

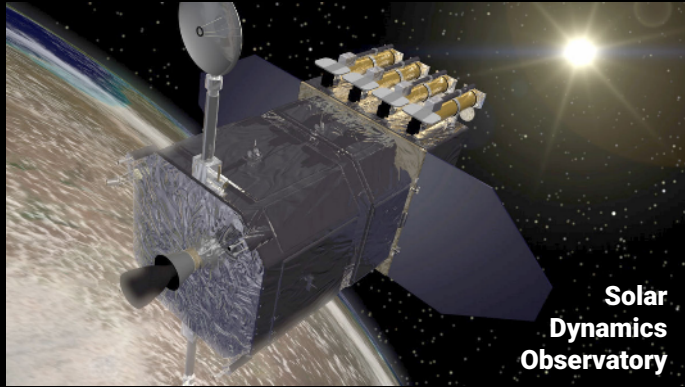


Can we use data-driven AI techniques to “revive” an instrument?

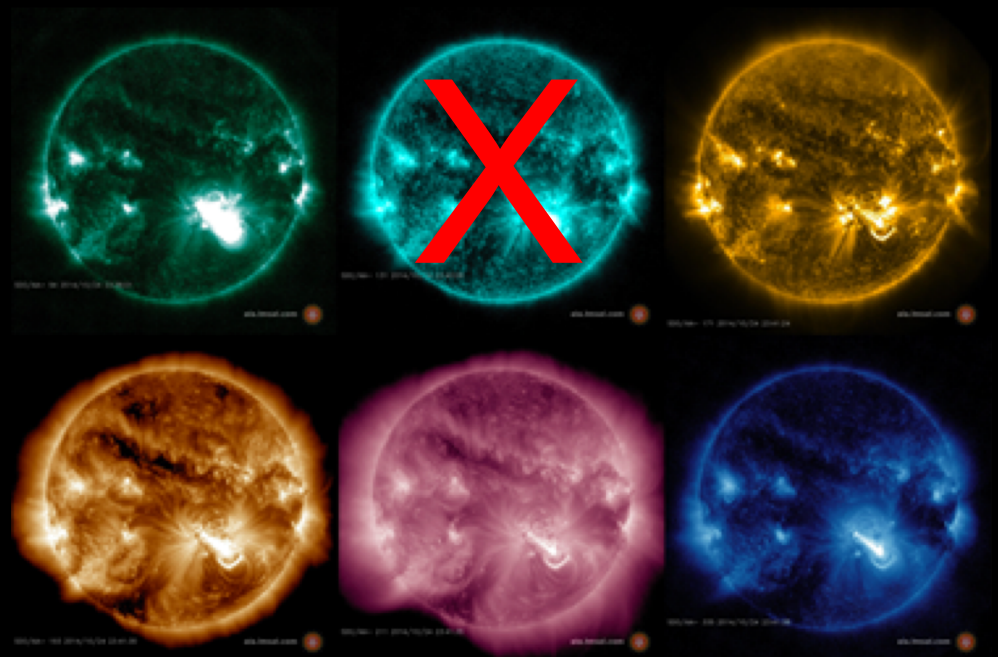
Richard Galvez, Rajat Thomas, Paul Wright,
Alexander Szenicer



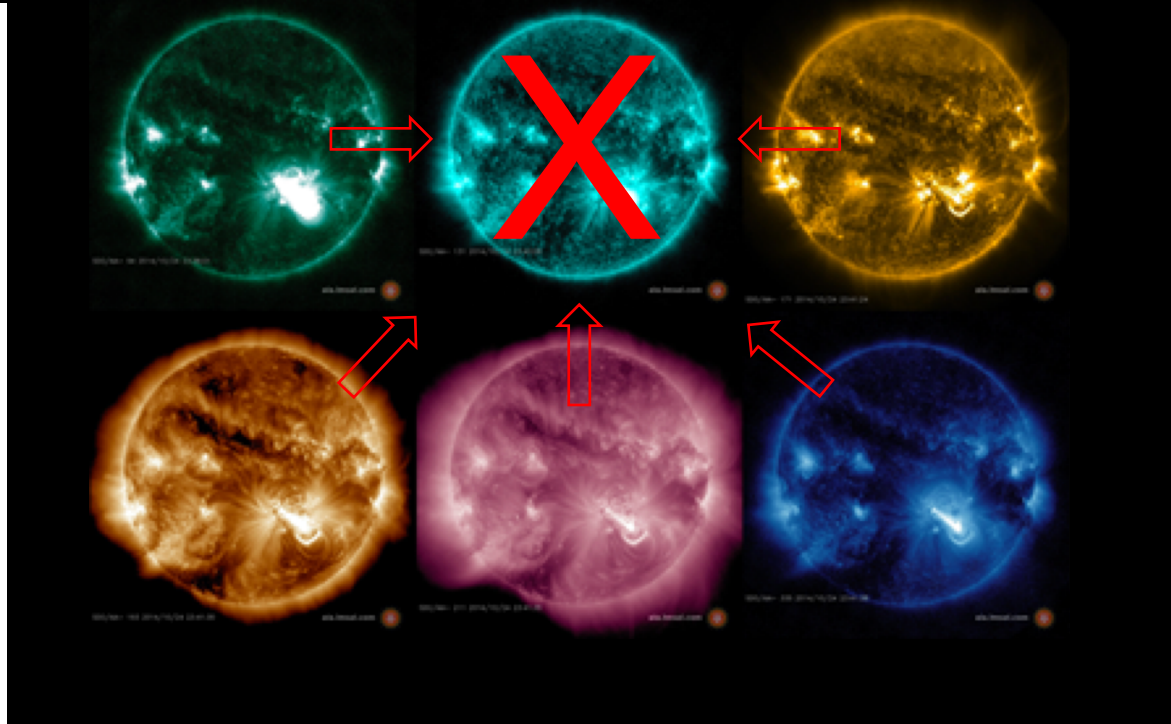
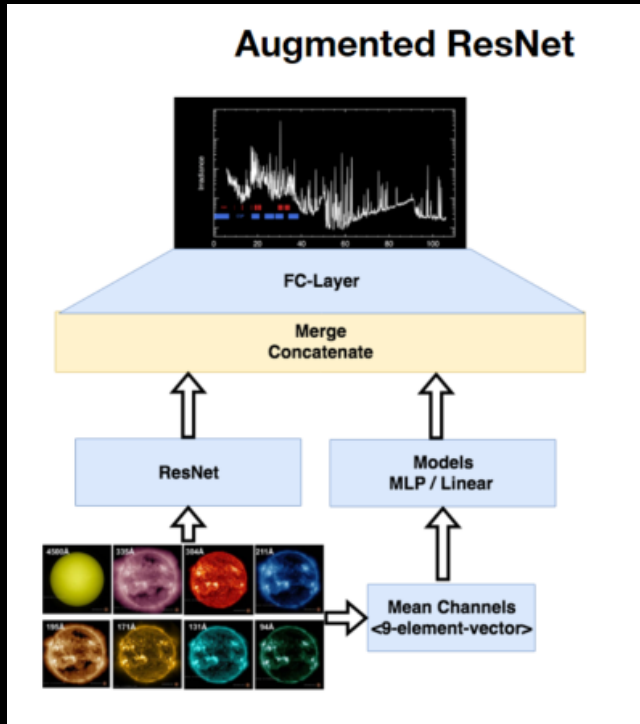
The Problem



The Problem

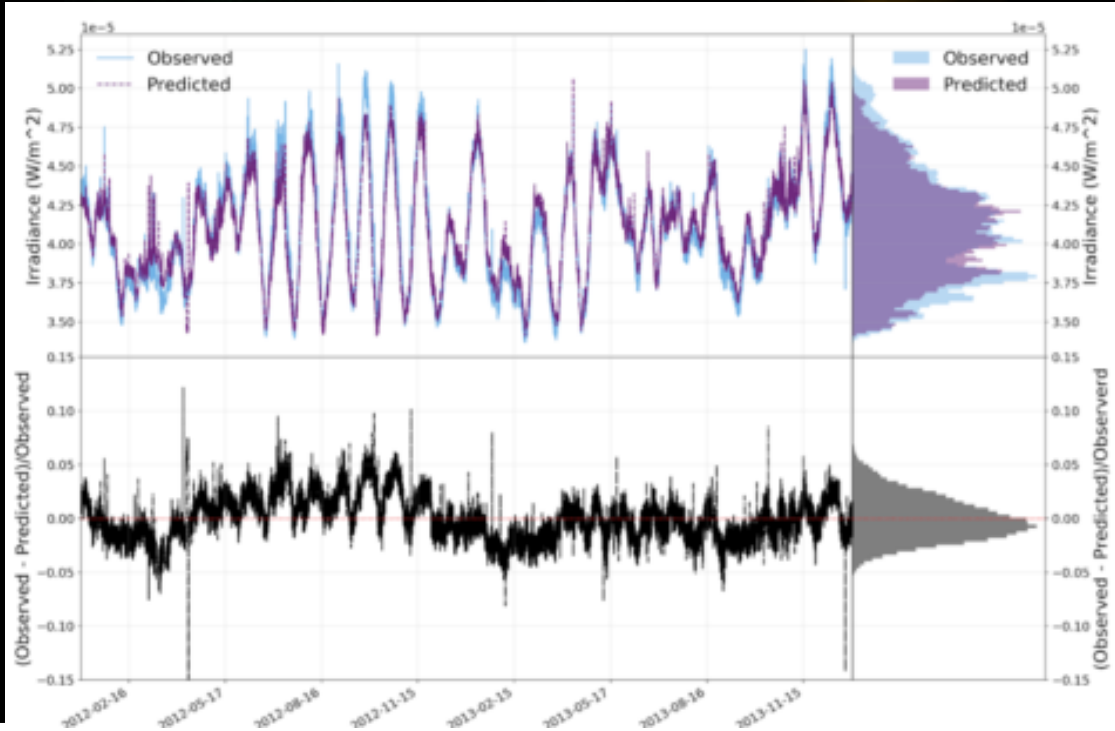
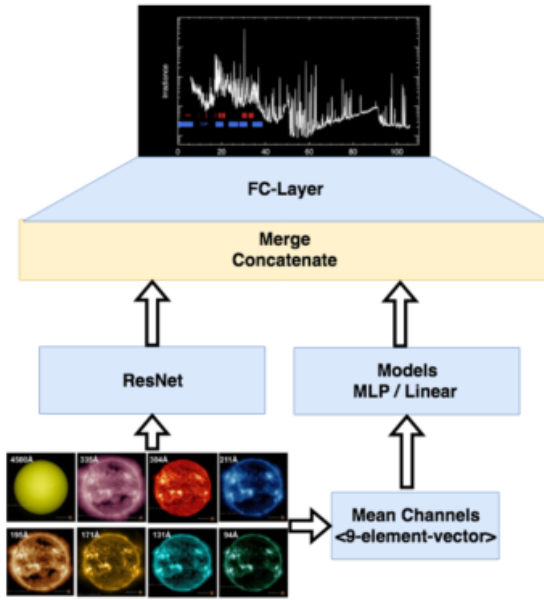


The Breakthrough



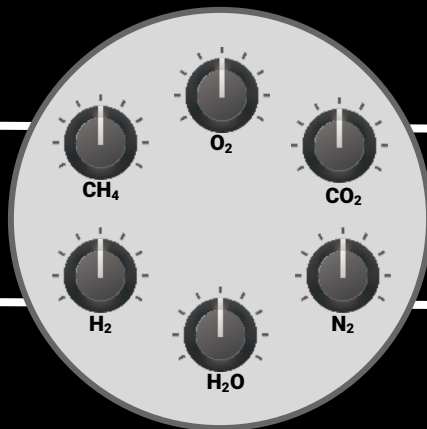
The Breakthrough

Augmented ResNet





**Generalized
Biology**



**Atmosphere
Dynamics**

Data Description

