Machine Learning for Exoplanet research

Lessons and results from
NASA Frontier Development Lab

Dr. Daniel Angerhausen
CSH Fellow
Bern University

NASA FDL

SETI INSTITUTE · intel AI · IBM · LOCKHEED MARTIN · SPACE RESOURCES.LU · XPRIZE · kx · Google Cloud · KBRwyle · NVIDIA

Sagan School 2009

Who am I?

UNIVERSITÄT BERN
CSH CENTER FOR SPACE AND HABITABILITY

explainables
SCIENCE COMMUNICATION

PlanetS

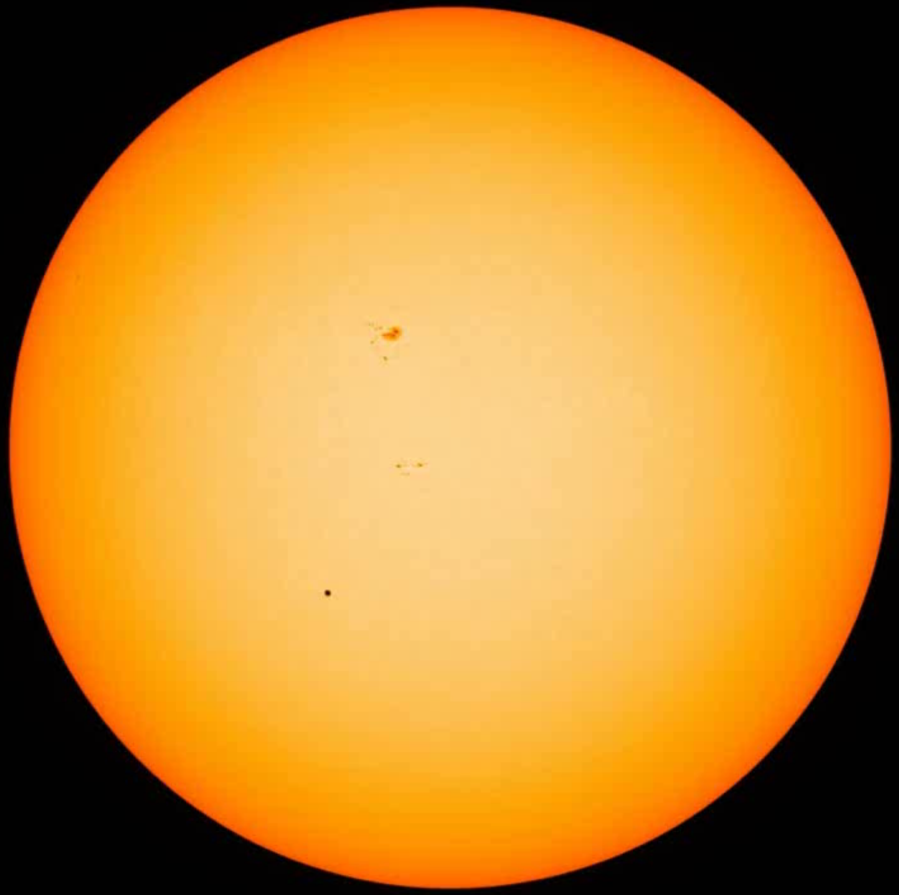
SETI INSTITUTE

Blue Marble Space Institute of Science
Celebrating 10 years of exploration

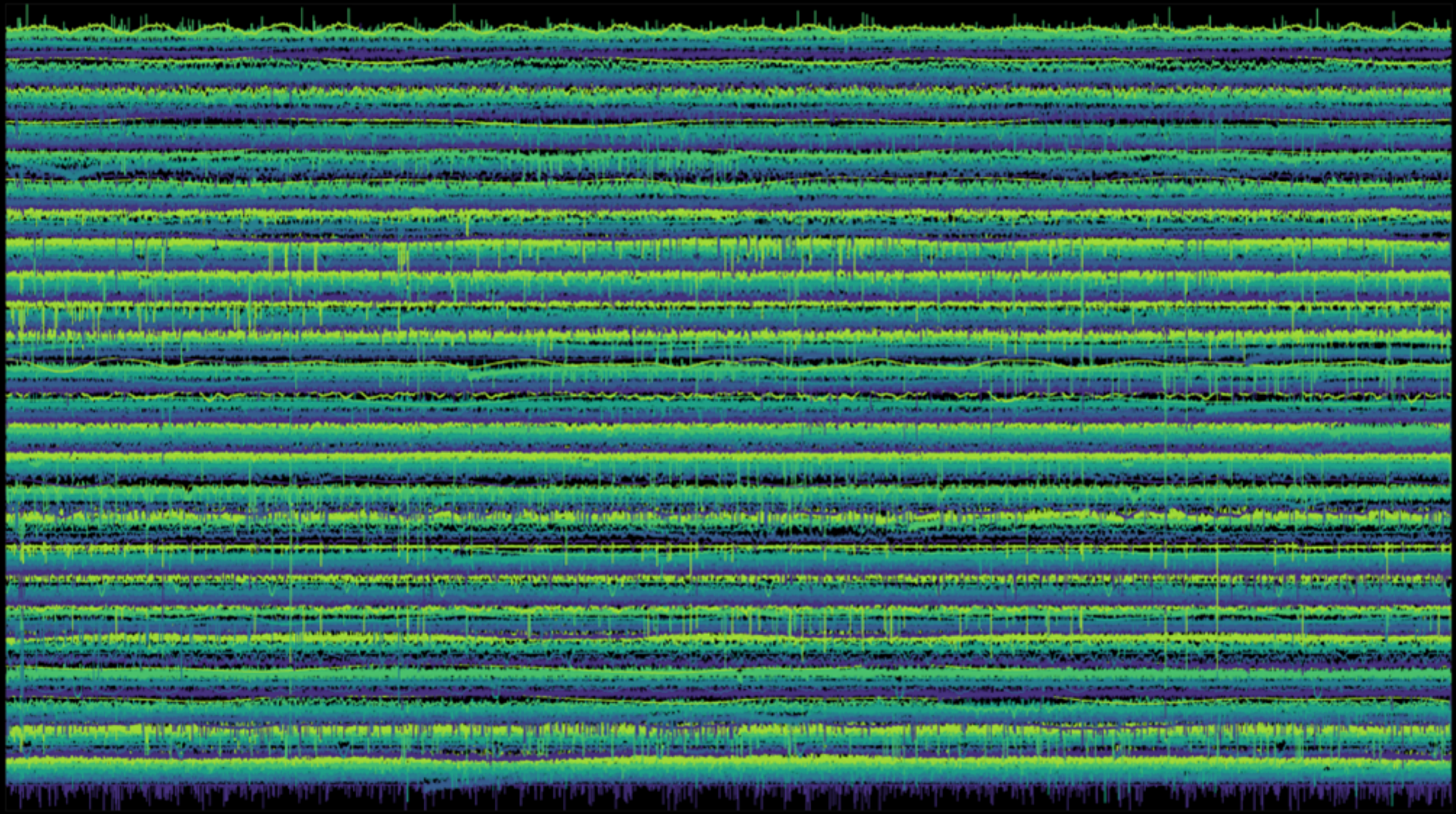
Google Cloud

NASA

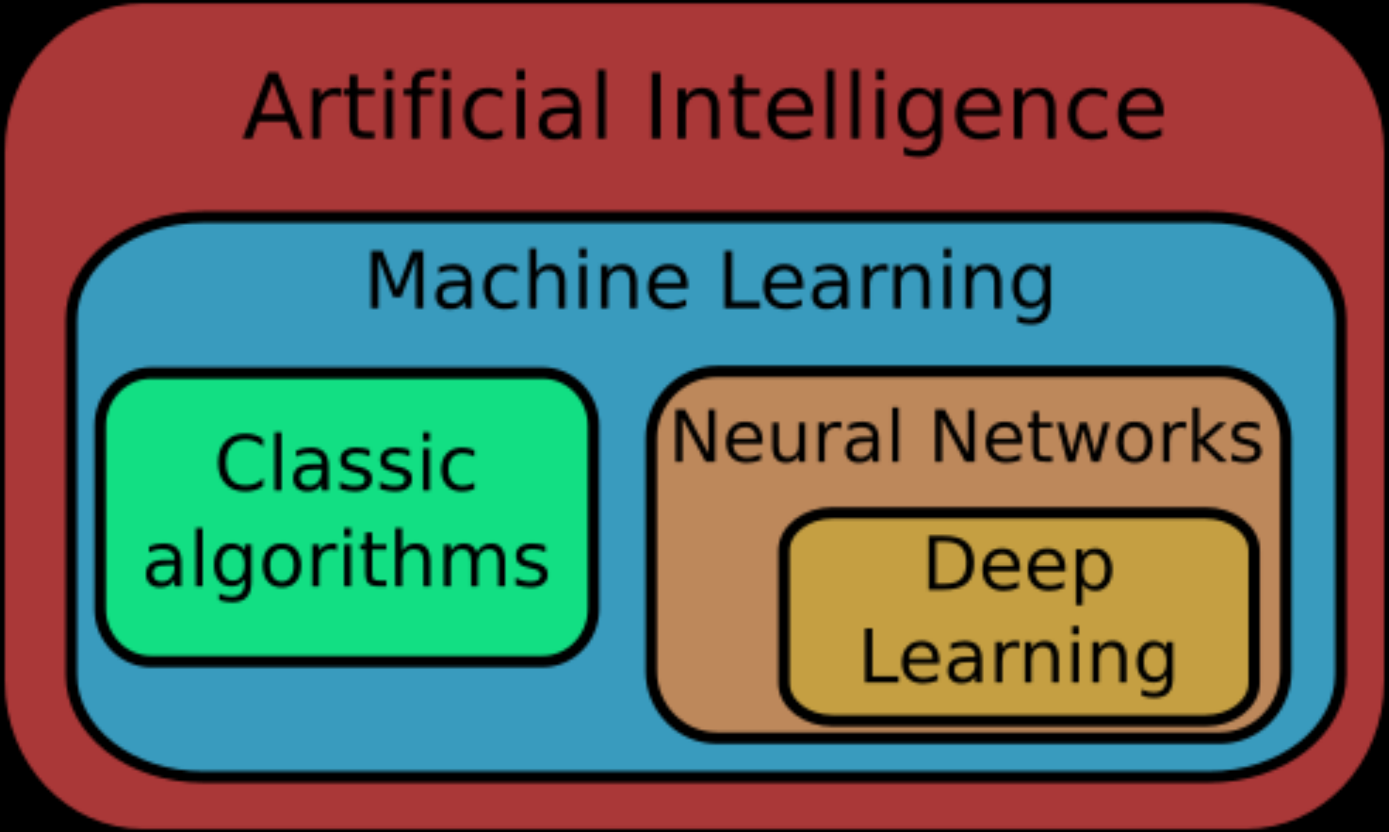
NASA FRONTIER DEVELOPMENT LAB



A mercury sized planet transiting a sun-like star

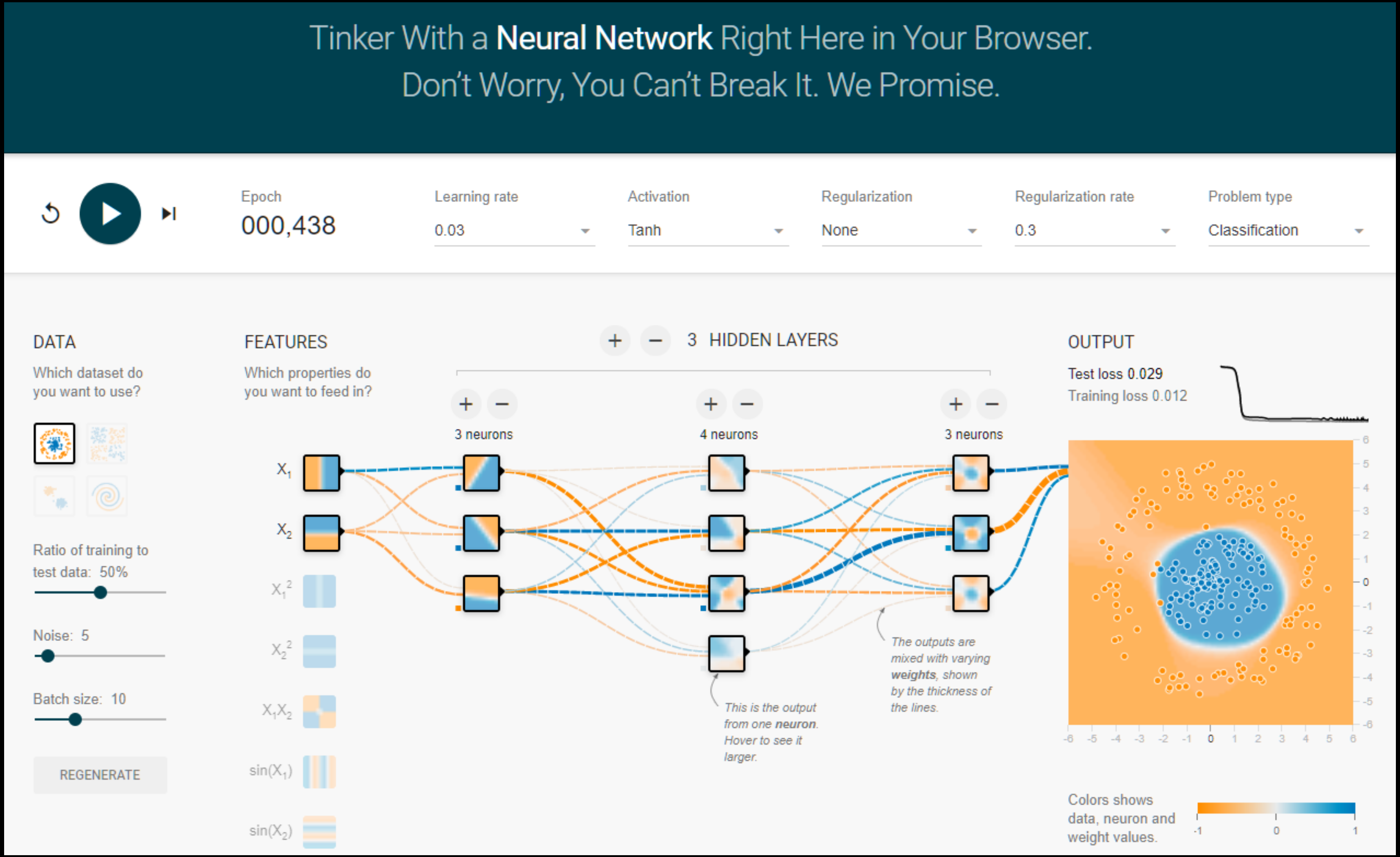


Data > Information > Knowledge > Wisdom


Artificial Intelligence
Machine Learning
Classic algorithms
Neural Networks
Deep Learning

Machine Learning
Classic algorithms
Deep learning
Reinforcement learning
Ensemble learning
Unsupervised learning
Supervised learning
Dimensionality reduction
PCA
NMF
GDA
LDA
Kernel PCA
Association rule learning
Eclat
FP-growth
Apriori
OPUS
Context bases
Clastering
Agglomerative clustering
K-means
Mean-shift
DBSCAN
Expectation-Minimization
Regression
Linear
Logistic
Lasso
Polinomial
Classification
Naive Bayes
Decision Trees
Random forest
Nearest Neighbour
Genetic algorithms
Monte Carlo
Q-Learning
SARSA
A3C
DQN
Bagging
Stacking
Boosting
Bucketing
Bayesian combination
Generative adversarial GAN
Perceptron
Convolutional CNN
Autoencoders
CAE
VAE
Sparse
Denoising
Reccurrent RNN
LSTM
GRU
Hopfield
IndRNN

playground.tensorflow.org



"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."

NASA FRONTIER DEVELOPMENT LAB

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 Soboczenski
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

PROBLEM PHASE

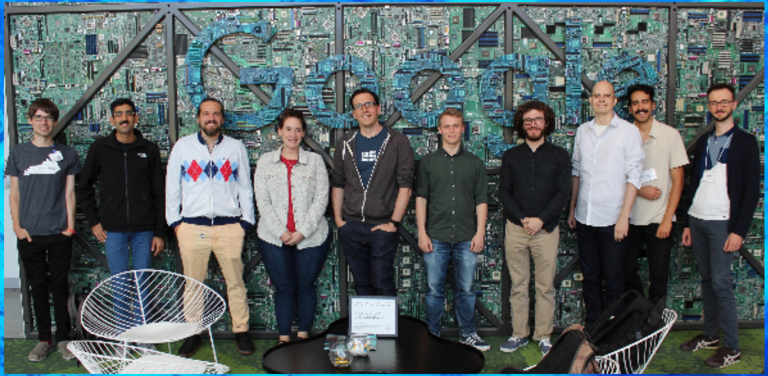
- **CONCEPT DEFINITION** 10th-14th June
- **PROTOTYPING** 26th-30th June
- **BIG IDEAS** 3rd-7th July

SOLUTION PHASE

- **DATA PREP & PROTOTYPING** 17th-21st July
- **PROTOTYPING & PIVOTING** 24th-28th July
- **PROTOTYPING & DEMO** 31st July - 4th Aug

DOCUMENT PHASE

- **DOCUMENT & DRAFT** 7th-11th Aug
- **PRESENTATIONS** 14th-18th Aug


Google


iamai


Carl Sagan Center
Carl Sagan Center

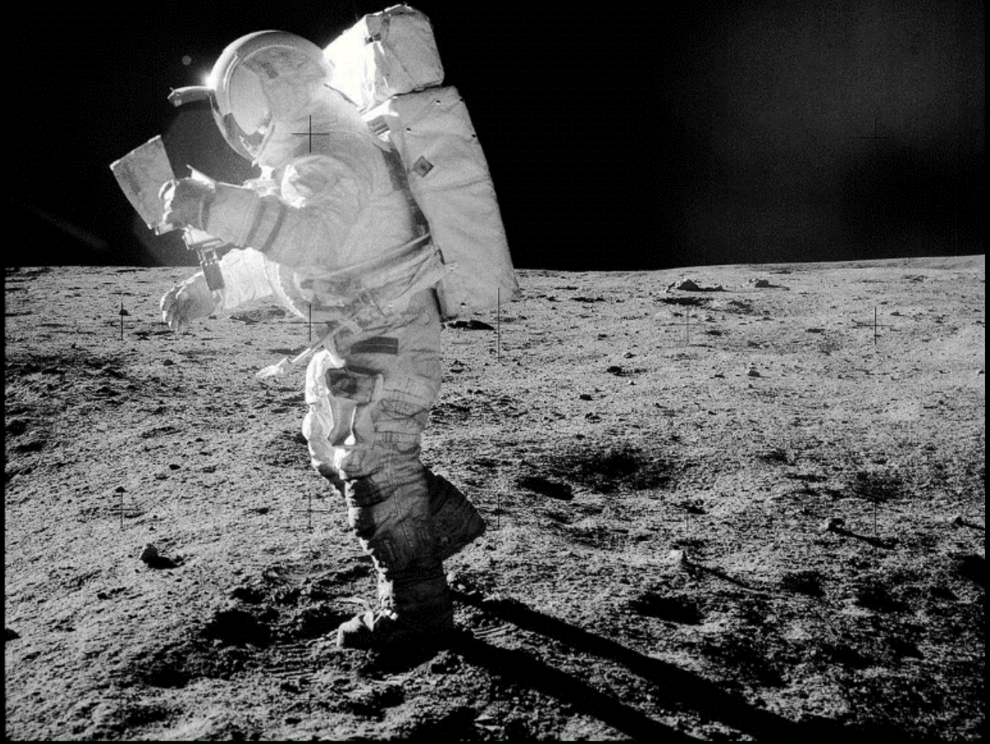

United States

Can we use AI techniques for localization on the Moon?

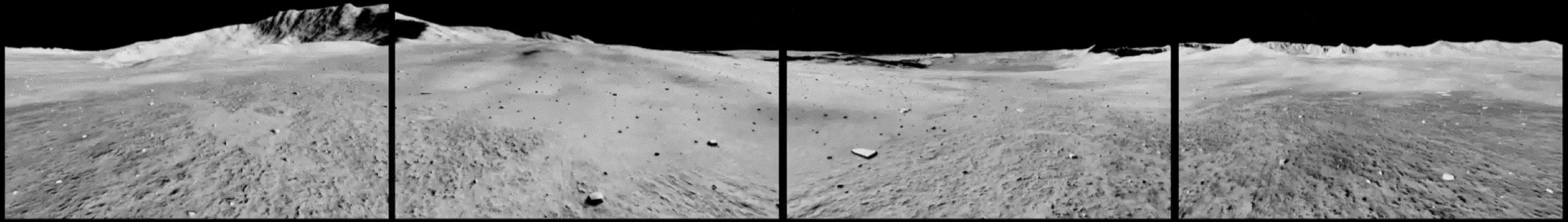
Andrew Chung, Philippe Ludivig, Ben Wu, Ross Potter

The Problem: Where are you?

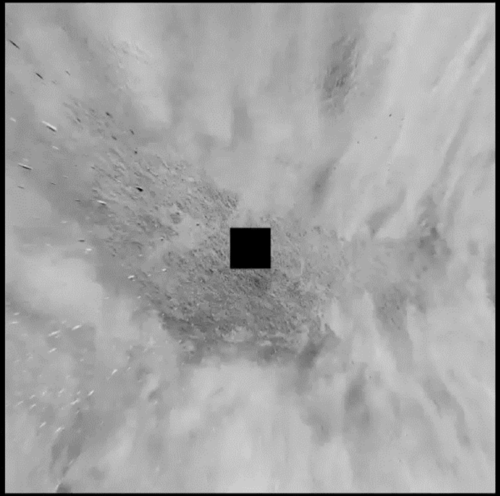
The Problem: Where are you?

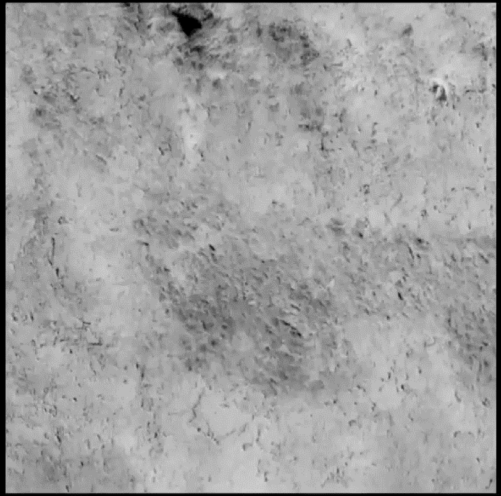


The Problem



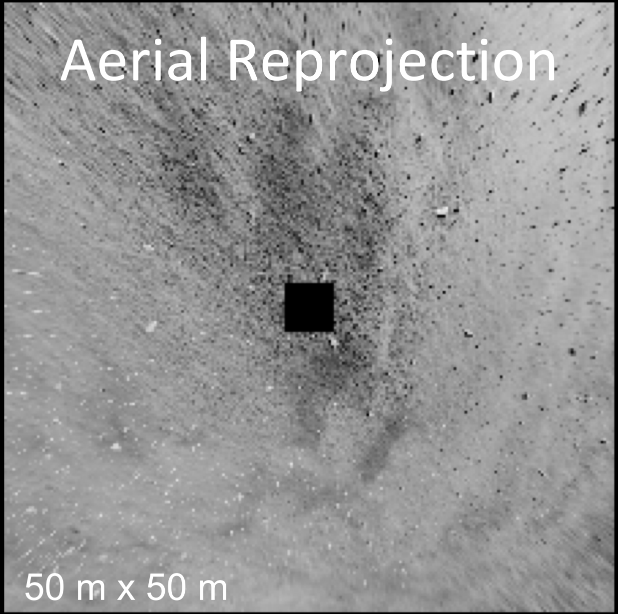
Reprojected View

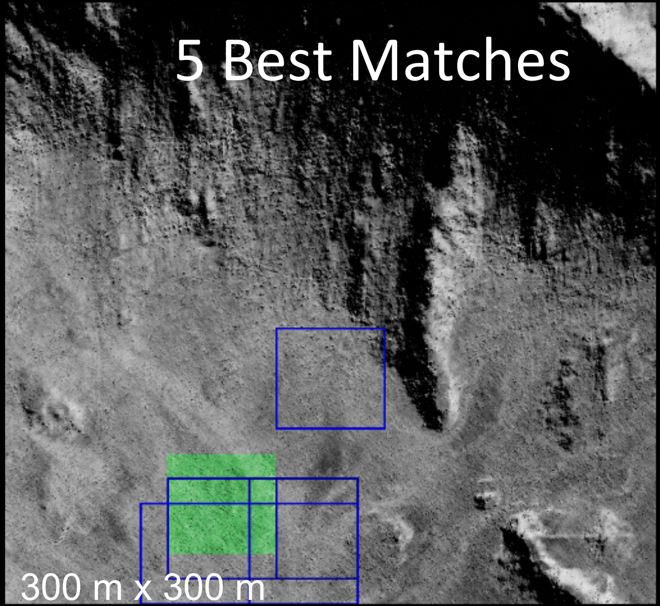




True Orbital View

The Breakthrough



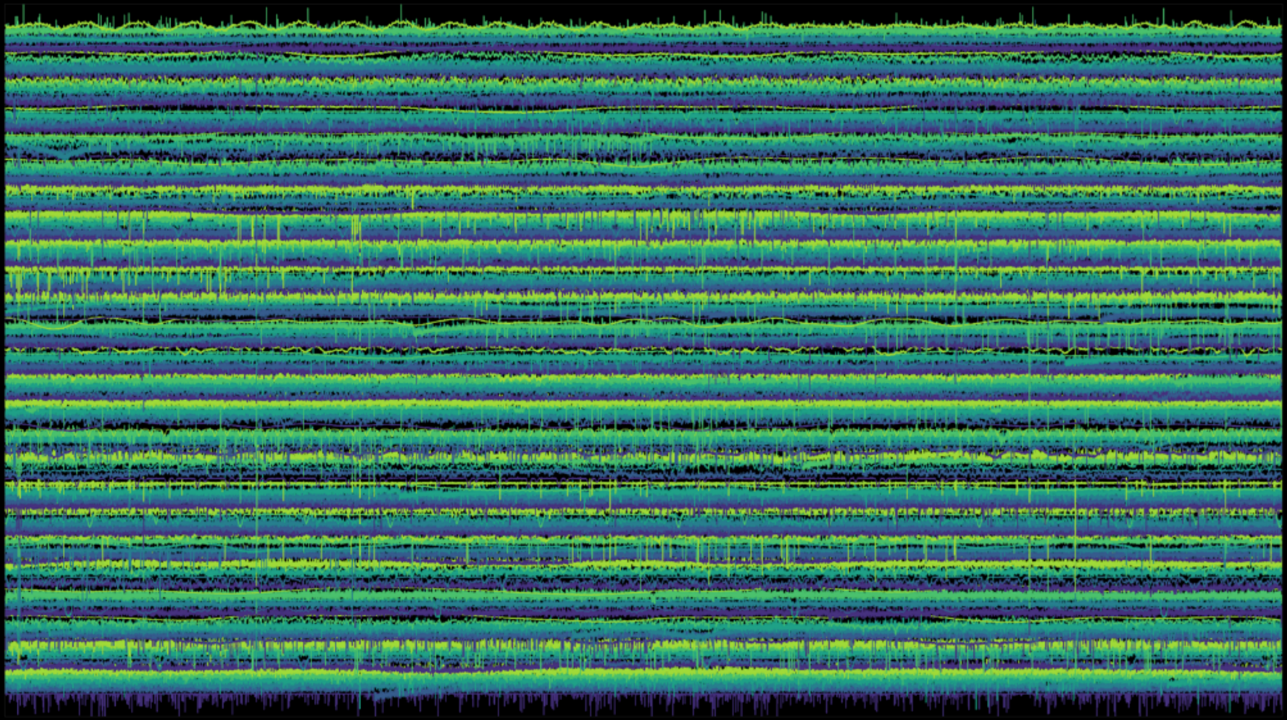


Increase the efficacy and yield of exoplanet transit detections with deep learning

Michele Sasdelli, Megan Ansdell, Hugh Osborn, Yani Ionnou

The Problem

Where are the planets and are they real?

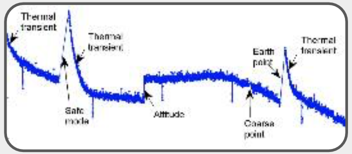


Kepler/TESS Pipelines

Target Pixel File (TPF)

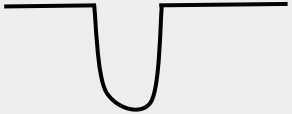
Smith+2012, Stumpe+2012

Aperture Photometry & Systematics Correction



Jenkins+2010, Seader+2013

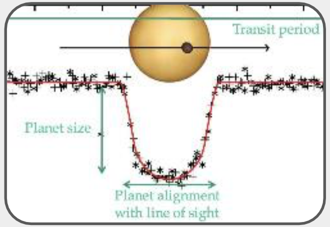
Transiting Planet Search (TPS)



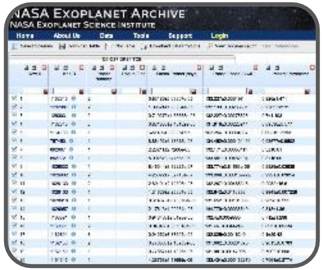
Threshold Crossing Event (TCE)

Wu+2010

Data Validation (DV)



Exoplanet Catalogues

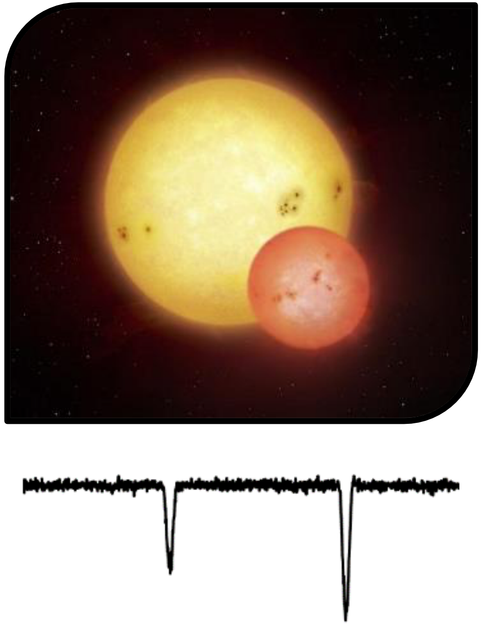


Batalha+2013, Burke+2014, Rowe+2015, Mullally+2015

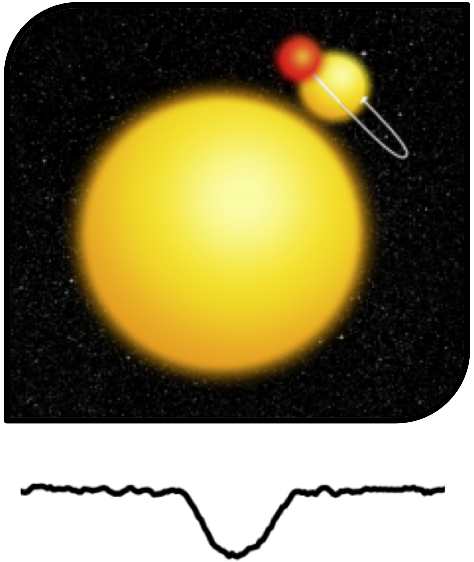
Kepler TCE Review Team [human vetting]

Where deep learning can (and is!) helping

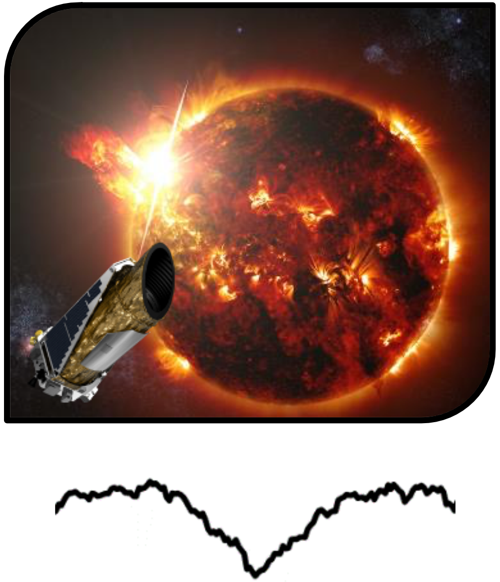
The Data: False Positives



Eclipsing Binaries (EBs)



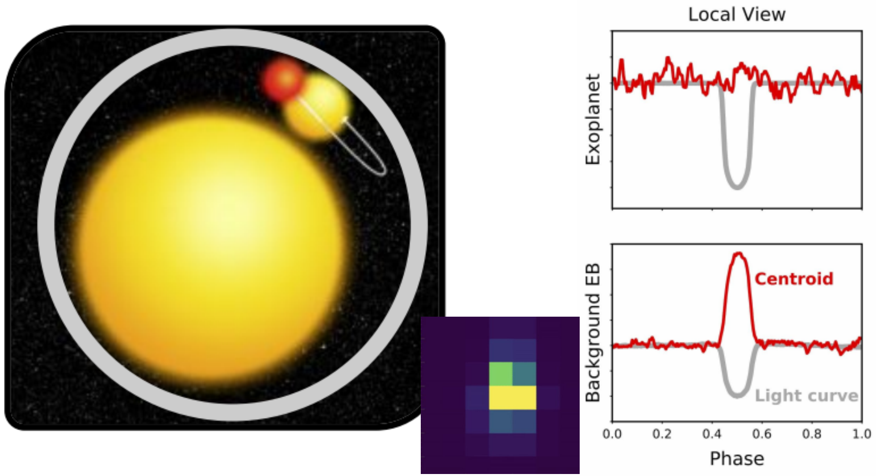
Background Eclipsing Binaries (BEBs)



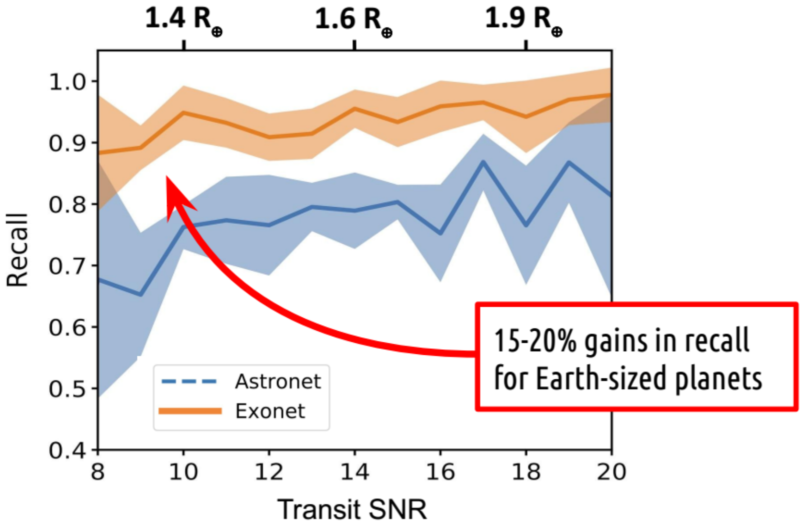
Stellar Variability / Instrumental Noise

Add domain knowledge

Improved model



Centroid shift



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 Soboczenski, Simone Zorzan, Molly O’Beirne

ASTROBIOLOGY

The Problem

N=1

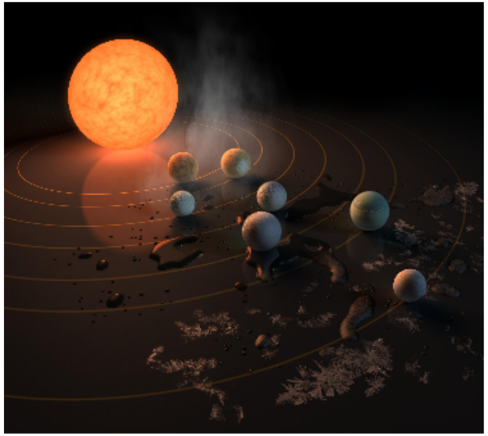
...not exactly BigData

Team 1

~200.000 ATMOSpheres Calculated

FDL

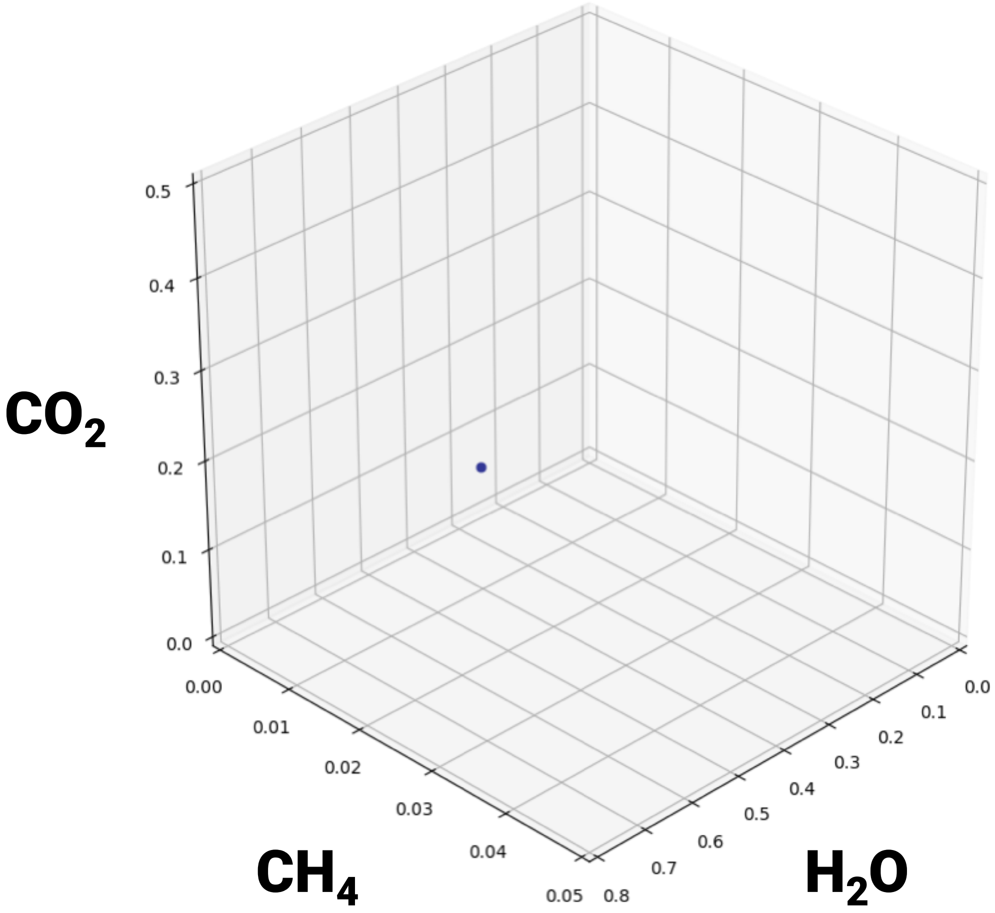
vpl **Atmos**



docker

pyATMOS

Google Cloud



0.5
0.4
0.3
0.2
0.1
0.0
CO_2
0.00
0.01
0.02
0.03
0.04
0.05
CH_4
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
H_2O

Team 2

3 Million

Observations of spectra simulated

FDL

Team 2

3 Million

Observations of spectra simulated

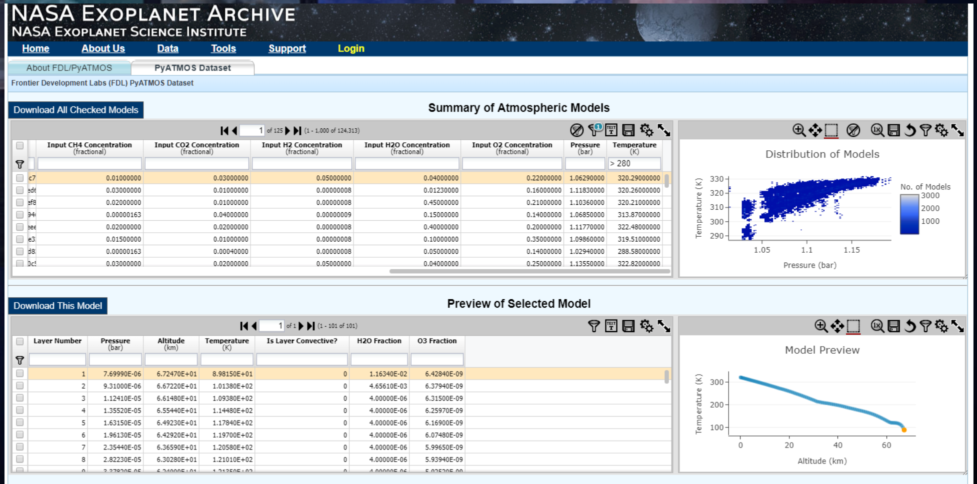
Google Cloud

The breakthrough

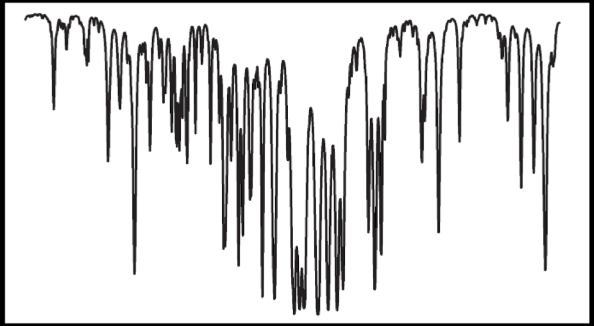
Datasets and software soon available:

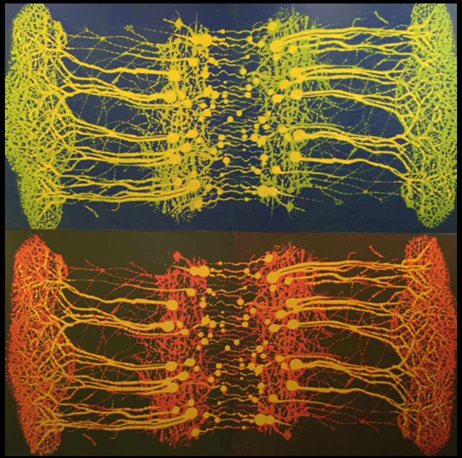
NASA Exoplanet Archive

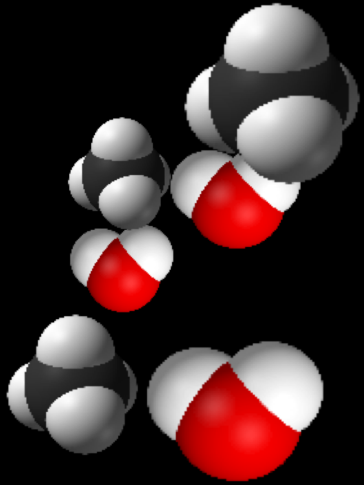
Google cloud/Kaggle



Spectrum > DL > Composition







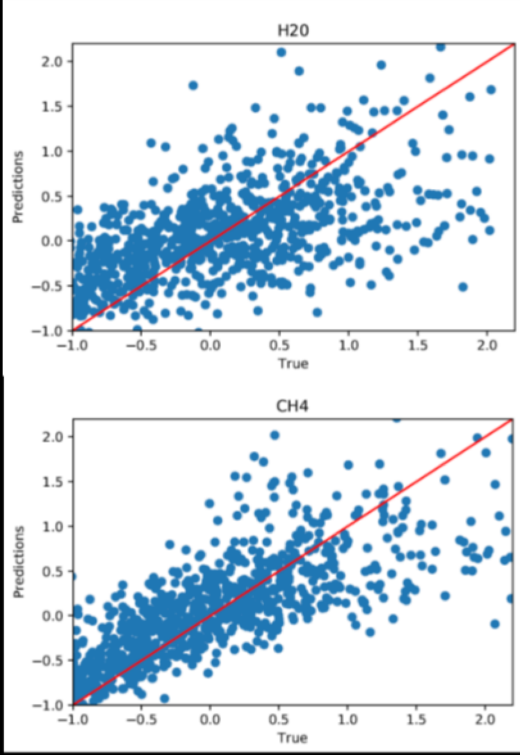
First application of (Emsembles of Bayesian)

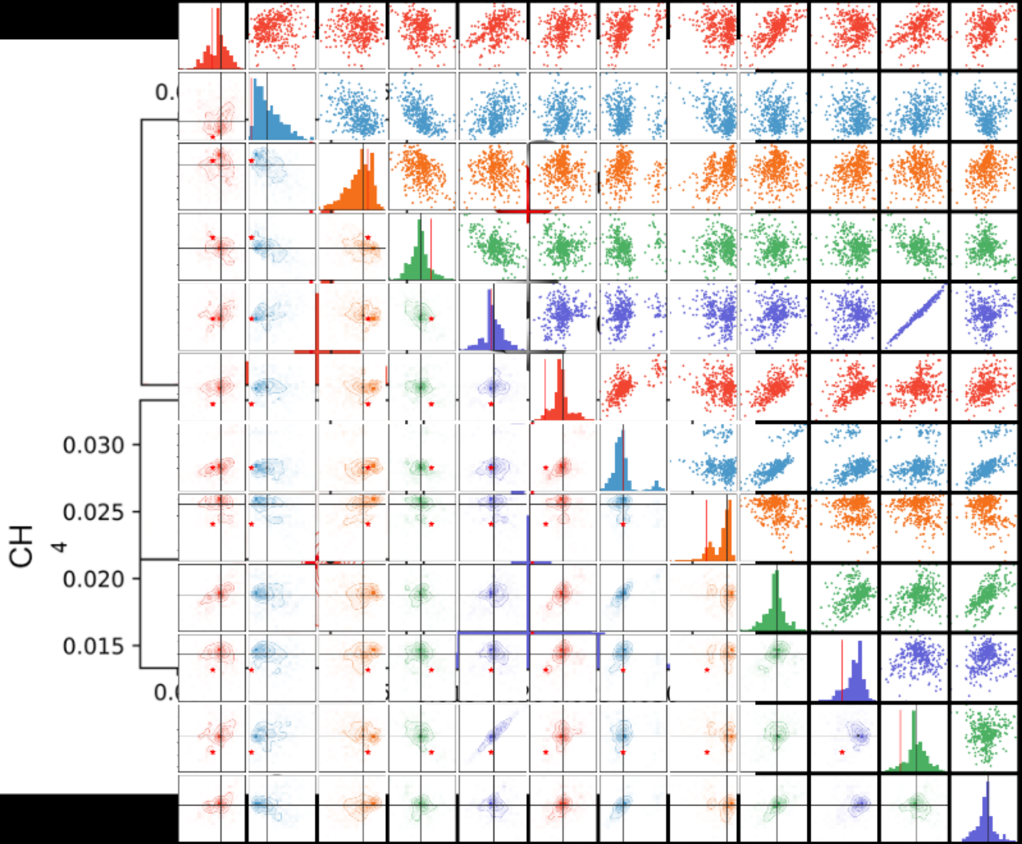
Neural Networks in Exoplanet Spectral Retrieval

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

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

Yes, but....


H20
Predictions
True
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
CH4
Predictions
True


CH
4
0.030
0.025
0.020
0.015

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)

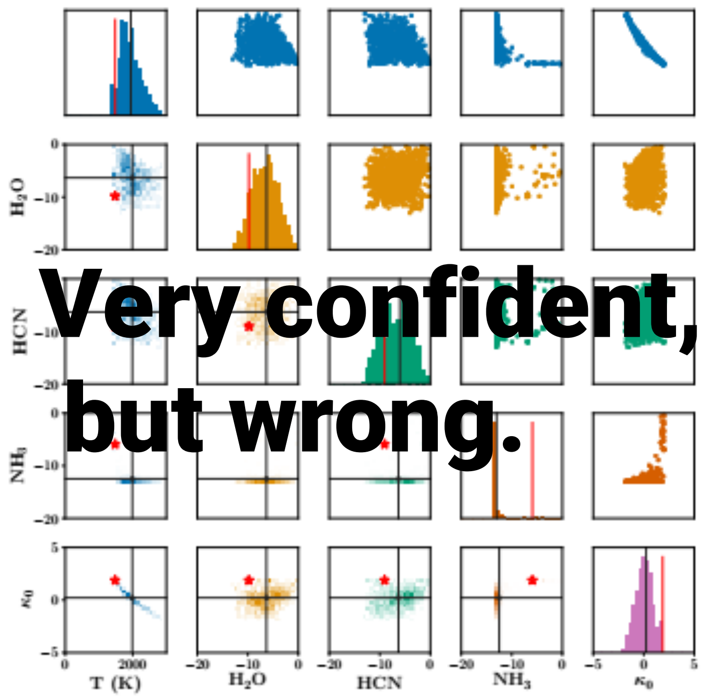
Comparison

WFC3 spectrum of WASP-12b

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

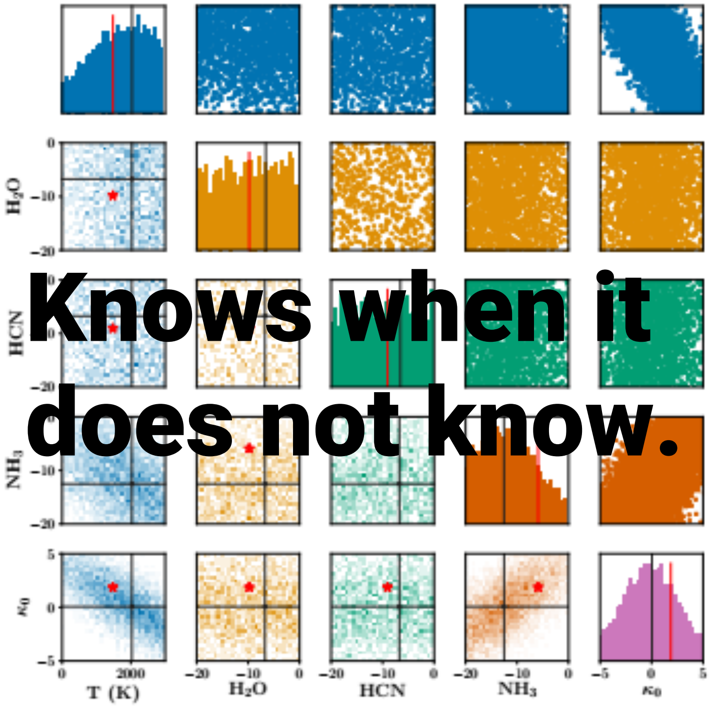
	T(K)	$\log X_{H_2O}$	$\log X_{HCN}$	$\log X_{NH_3}$	κ_0	Mean
plan-net R^2	**0.770**	0.623	0.487	0.721	0.750	**0.673**
Ens. 5 plan-net R^2	**0.770**	**0.629**	**0.491**	**0.723**	**0.751**	**0.673**
Our Ran. Forest R^2	0.746	0.608	0.466	0.700	0.736	0.651
Ran. Forest[a] R^2	0.746	0.608	0.467	0.700	0.737	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

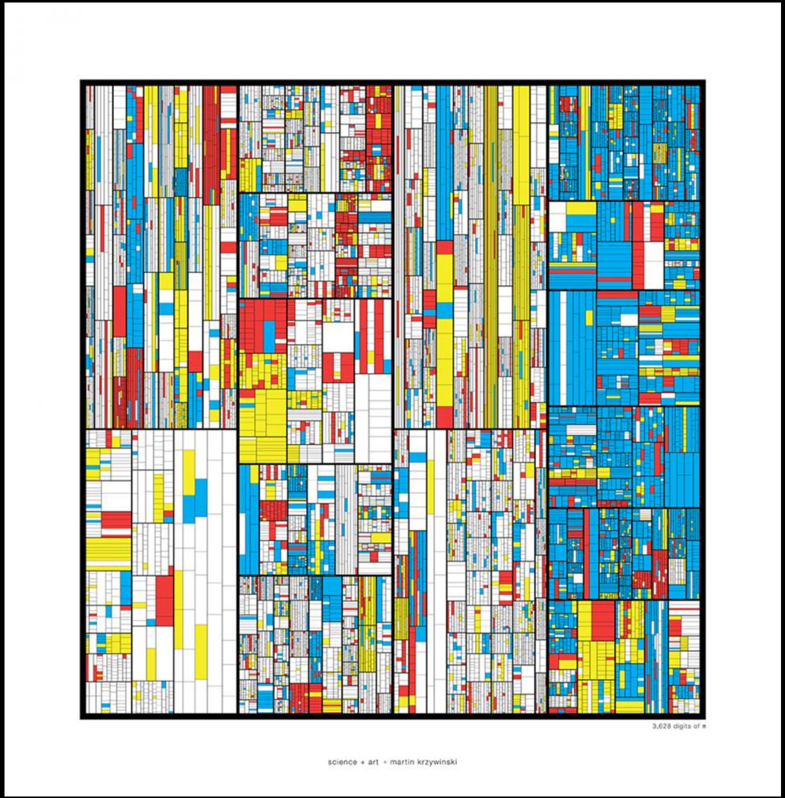
-You are already doing it

-Data is the driver

-AI/ML a toolbox, not one hammer

- DL new tool with some applications

-ask an Expert, collaborate



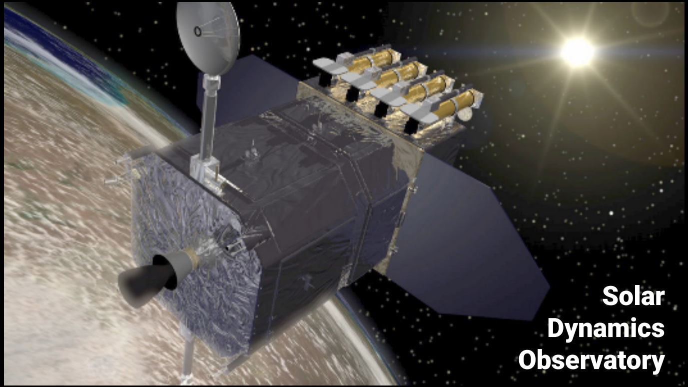

~~When~~ If we find the *first signs of life in space*:
… machine learning was used
… within a public private partnership

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

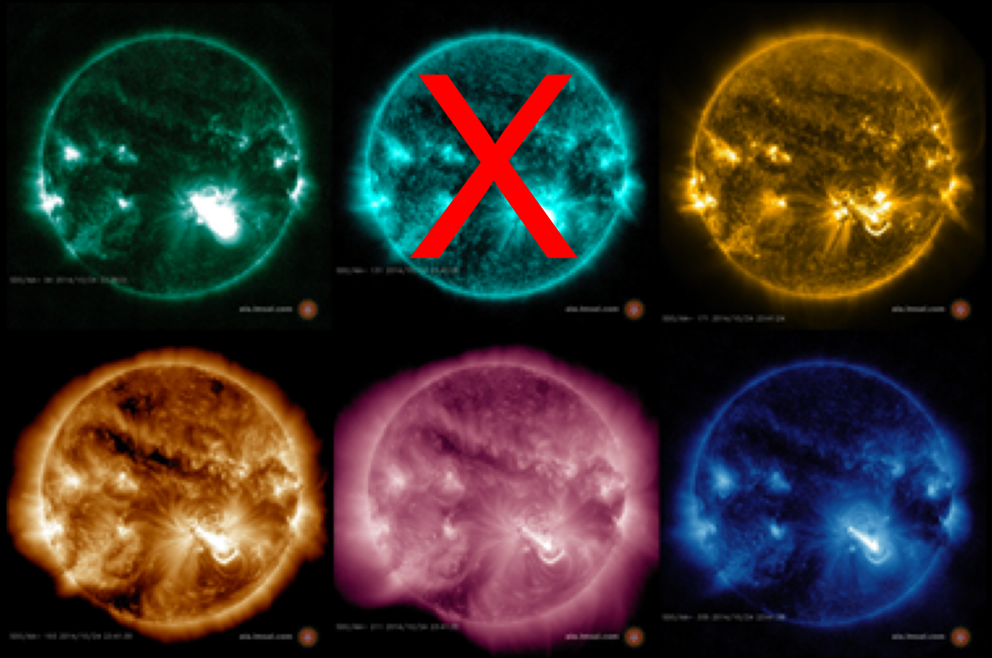
Richard Galvez, Rajat Thomas, Paul Wright, Alexander Szenicer

SPACE WEATHER

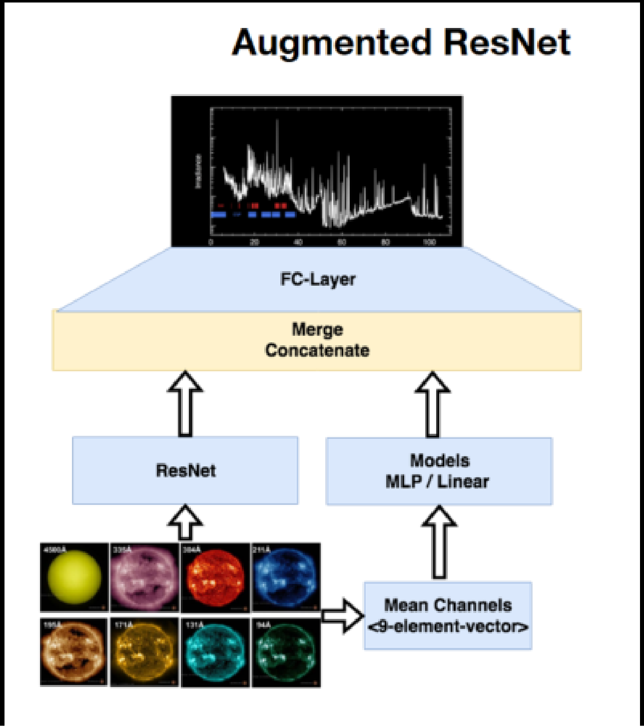
The Problem

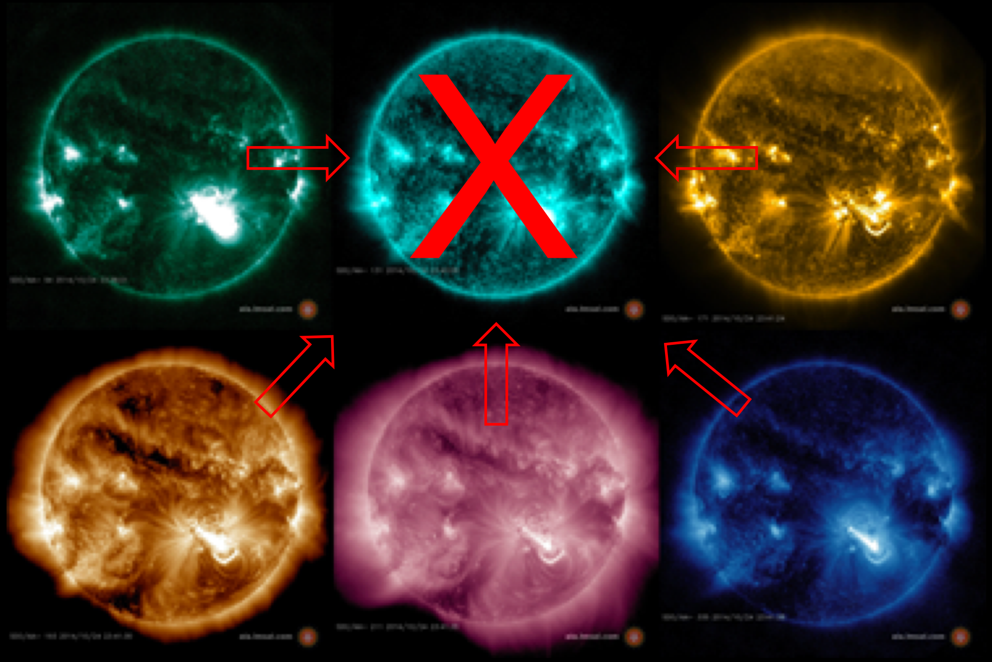


The Problem

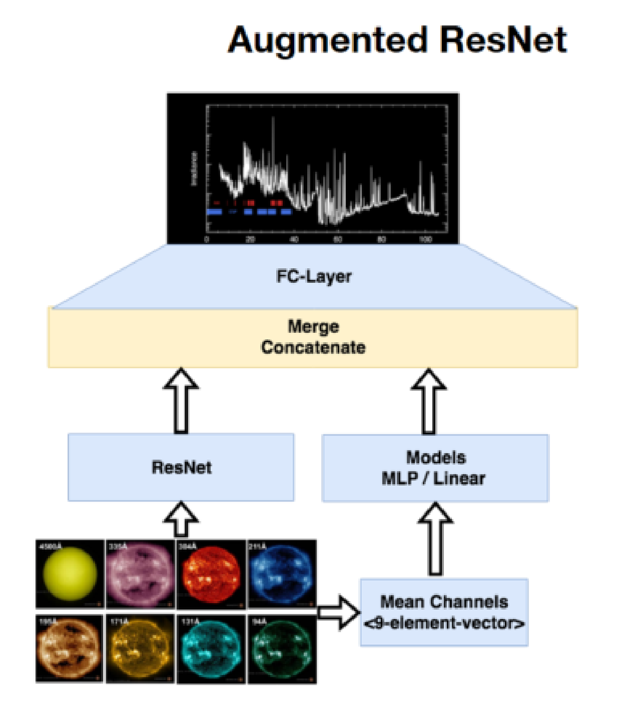


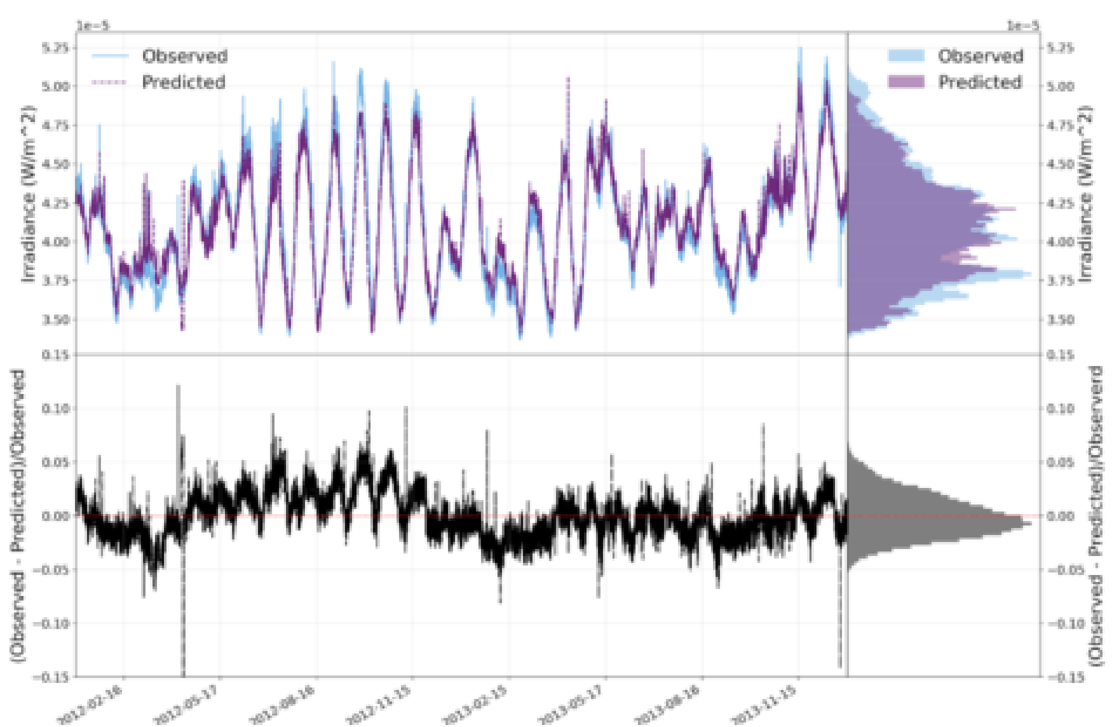
The Breakthrough

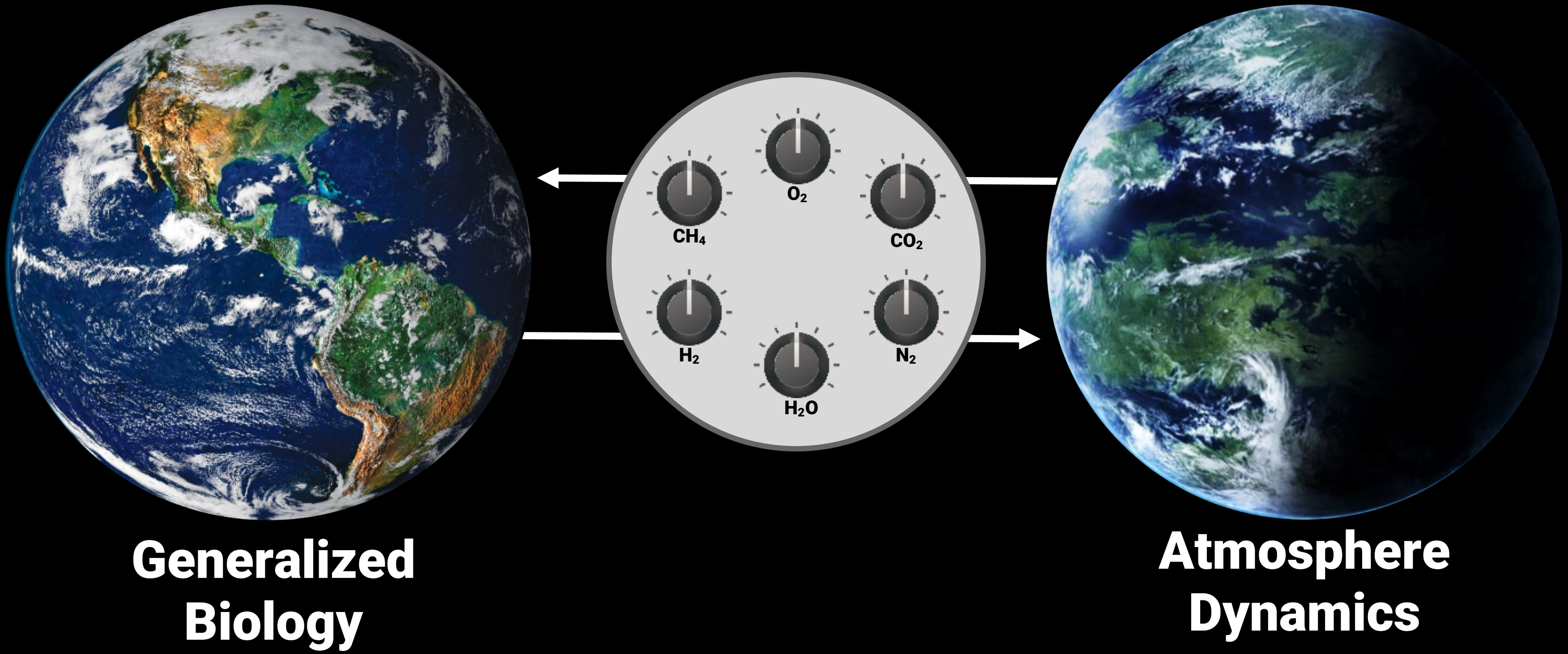




The Breakthrough



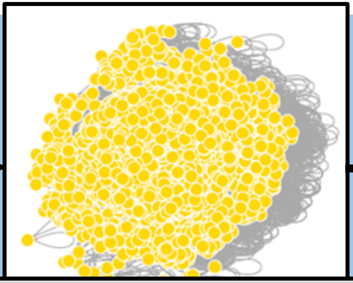




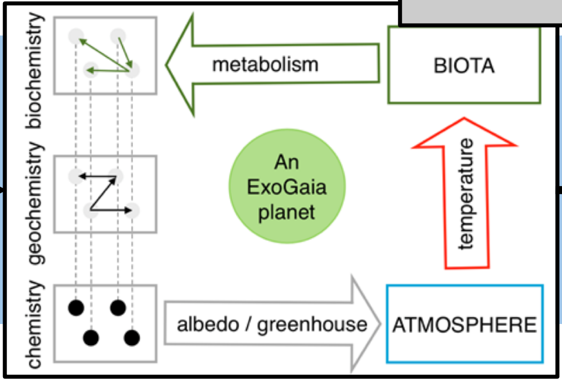
O_2
CH_4
CO_2
H_2
N_2
H_2O
Generalized Biology
Atmosphere Dynamics

Data Description

KEGG
[Real World]

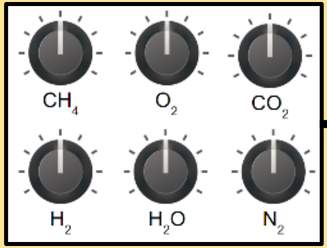


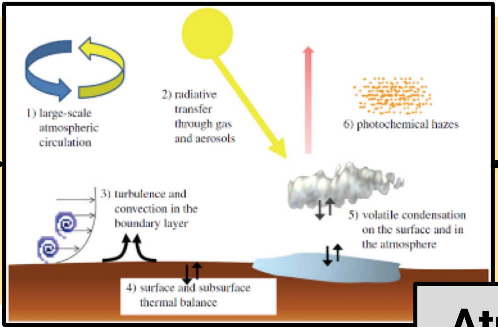
Biological Network



Abstract Exo-Gaia

Realistic Biological Feedback





Atmos

Atmosphere Climate Profile