2025 Exoplanet Research

[Christina Liu (Lakeside School, Seattle, WA)]

Overview of My Research

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost
  - Conducting **Feature Importance Analysis**

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost
  - Conducting **Feature Importance Analysis**
    - **Sh**apley **A**dditive Ex**p**lanations (SHAP)

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost
  - Conducting **Feature Importance Analysis**
    - **Sh**apley **A**dditive Ex**p**lanations (SHAP)

- **Exoplanet Classification**

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost
  - Conducting **Feature Importance Analysis**
    - **Sh**apley **A**dditive Ex**p**lanations (SHAP)
- **Exoplanet Classification**
  - Updating S. Seager's 2013 paper – *Exoplanet Habitability*

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost
  - Conducting **Feature Importance Analysis**
    - **Sh**apley **A**dditive Ex**p**lanations (SHAP)

- **Exoplanet Classification**
  - Updating S. Seager's 2013 paper – *Exoplanet Habitability*
  - Planetary System Architecture

Overview of My Research

- **Exoplanet Habitability** Through A **Machine Learning Perspective**
  - **Building ML Models** with High Precision
    - Random Forest
    - XGBoost
  - Conducting **Feature Importance Analysis**
    - **Sh**apley **A**dditive Ex**p**lanations (SHAP)

- **Exoplanet Classification**
  - Updating S. Seager's 2013 paper – *Exoplanet Habitability*
  - Planetary System Architecture
  - Planetary Formation (Literary Studies)

Exoplanet **Habitability**

Exoplanet **Habitability**

Exoplanet **Habitability**

Are there any other worlds where we can live?

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**
  - Presence of liquid water

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**
  - Presence of liquid water
  - Planetary surface temperature

Exoplanet **Habitability**

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**
  - Presence of liquid water
  - Planetary surface temperature
  - Planet atmosphere

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**
  - Presence of liquid water
  - Planetary surface temperature
  - Planet atmosphere
  - Planetary orbit

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**
  - Presence of liquid water
  - Planetary surface temperature
  - Planet atmosphere
  - Planetary orbit

- **Stellar** features + **stellar system** architecture

Exoplanet **Habitability**

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- Planet is within a **habitable zone**
  - Presence of liquid water
  - Planetary surface temperature
  - Planet atmosphere
  - Planetary orbit

- **Stellar** features + **stellar system** architecture
  - Stellar luminosity, variability, activity, metallicity, etc.

Exoplanet Habitability

Are there any other worlds where we can live?

KEY THINGS TO LOOK FOR/AT:

- ✦ Planet is within a **habitable zone**
  - ✧ Presence of liquid water
  - ✧ Planetary surface temperature
  - ✧ Planet atmosphere
  - ✧ Planetary orbit

- ✦ **Stellar** features + **stellar system** architecture
  - ✧ Stellar luminosity, variability, activity, metallicity, etc.
  - ✧ Other exoplanets in the system; impacts

Exoplanet Habitability

ML is usually used in the astrophysics community to determine whether a new exoplanet should be considered habitable through comparisons to existing planets with known categorizations.

Exoplanet Habitability

ML is usually used in the astrophysics community to determine whether a new exoplanet should be considered habitable through comparisons to existing planets with known categorizations.

My research instead focuses on the individual features of exoplanets, and how much they impact habitability → **feature importance analysis.**

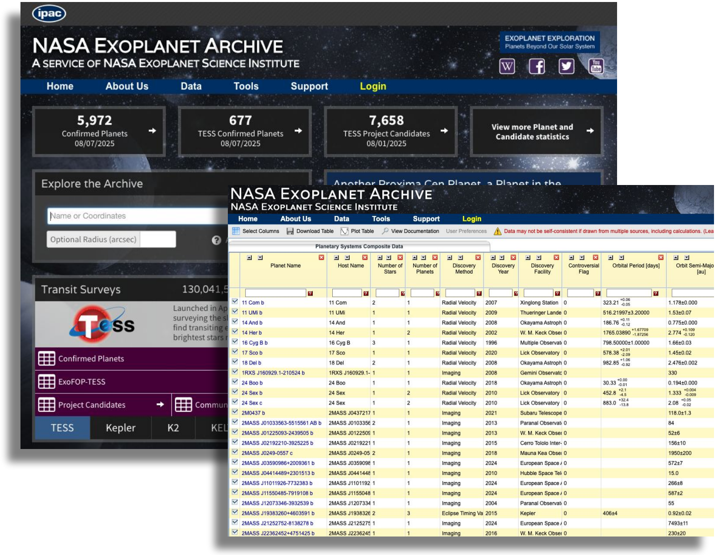
Exoplanet Habitability

Where do we get the features?

Exoplanet Habitability

Where do we get the features?

NASA Exoplanet Archive

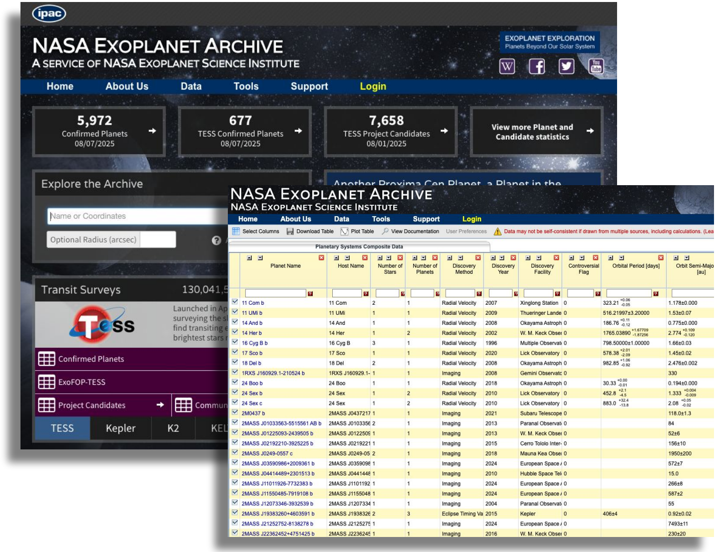


- ✦ Planetary Systems Composite Data

Exoplanet Habitability

Where do we get the features?

NASA Exoplanet Archive

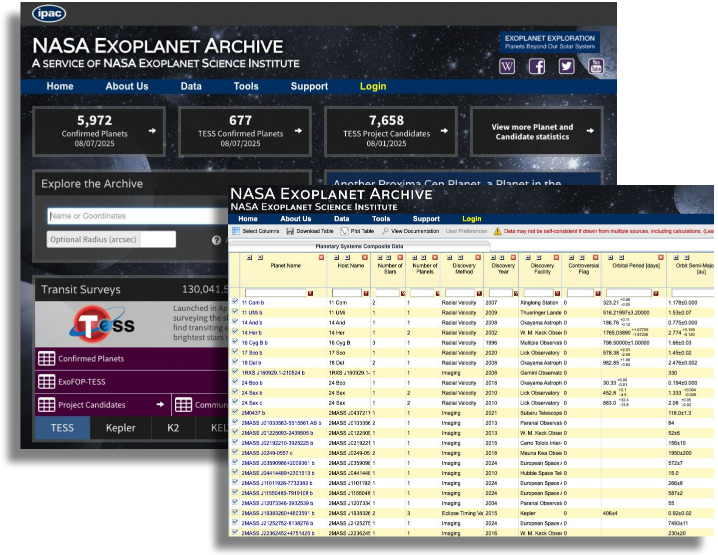


- ✦ Planetary Systems Composite Data
  - ✧ 5,834 confirmed exoplanets

Exoplanet Habitability

Where do we get the features?

NASA Exoplanet Archive

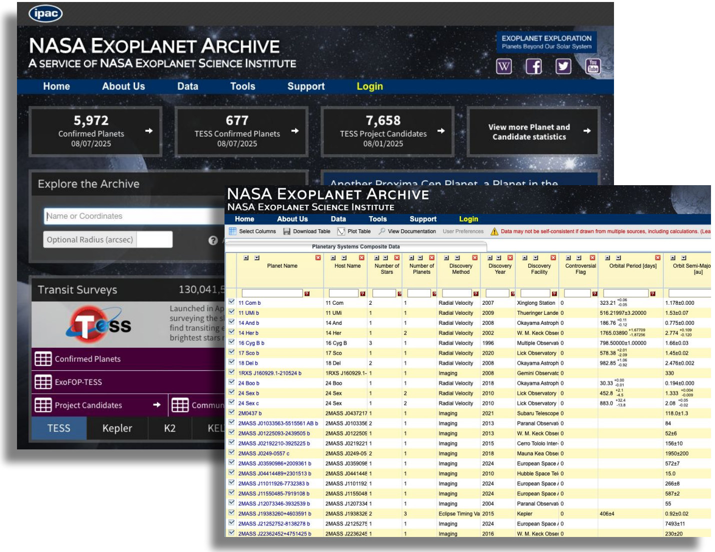


- ✦ Planetary Systems Composite Data
  - ✧ 5,834 confirmed exoplanets
  - ✧ Features including planetary orbit semi-major axis, planet radius, planet mass, stellar effective temperature, stellar mass, etc.

Exoplanet Habitability

Where do we get the features?

NASA Exoplanet Archive



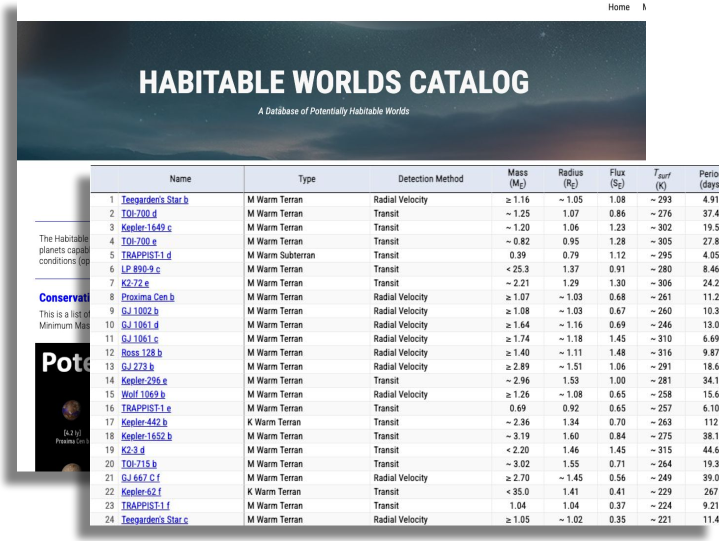
- ✦ Planetary Systems Composite Data
  - ✧ 5,834 confirmed exoplanets
  - ✧ Features including planetary orbit semi-major axis, planet radius, planet mass, stellar effective temperature, stellar mass, etc.

Missing a label to train our data.

Exoplanet Habitability

Where do we get the features?

Habitable Worlds Catalogue, PHL @ UPR Arecibo



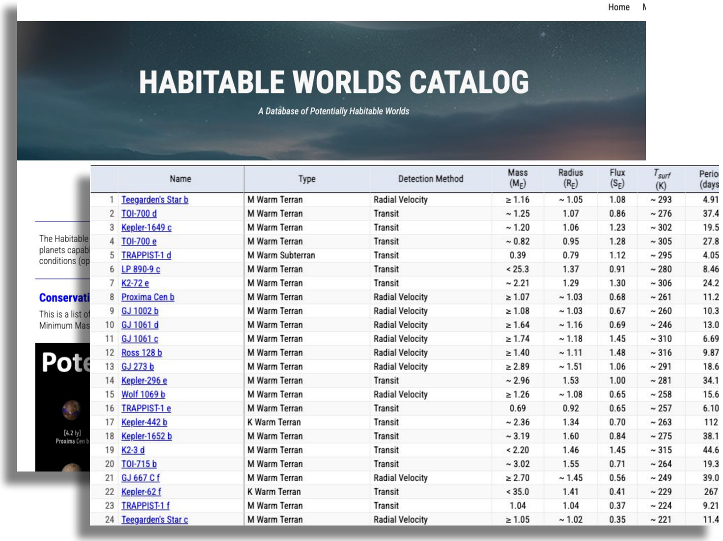
	Name	Type	Detection Method	Mass (M_E)	Radius (R_E)	Flux (S_E)	T_{surf} (K)	Perio (days
1	Teegarden's Star b	M Warm Terran	Radial Velocity	≥ 1.16	~ 1.05	1.08	~ 293	4.91
2	TOI-700 d	M Warm Terran	Transit	~ 1.25	1.07	0.86	~ 276	37.4
3	Kepler-1649 c	M Warm Terran	Transit	~ 1.20	1.06	1.23	~ 302	19.5
4	TOI-700 e	M Warm Terran	Transit	~ 0.82	0.95	1.28	~ 305	27.8
5	TRAPPIST-1 d	M Warm Subterran	Transit	0.39	0.79	1.12	~ 295	4.05
6	LP 890-9 c	M Warm Terran	Transit	< 25.3	1.37	0.91	~ 280	8.46
7	K2-72 e	M Warm Terran	Transit	~ 2.21	1.29	1.30	~ 306	24.2
8	Proxima Cen b	M Warm Terran	Radial Velocity	≥ 1.07	~ 1.03	0.68	~ 261	11.2
9	GJ 1002 b	M Warm Terran	Radial Velocity	≥ 1.08	~ 1.03	0.67	~ 260	10.3
10	GJ 1061 d	M Warm Terran	Radial Velocity	≥ 1.64	~ 1.16	0.69	~ 246	13.0
11	GJ 1061 c	M Warm Terran	Radial Velocity	≥ 1.74	~ 1.18	1.45	~ 310	6.69
12	Ross 128 b	M Warm Terran	Radial Velocity	≥ 1.40	~ 1.11	1.48	~ 316	9.87
13	GJ 273 b	M Warm Terran	Radial Velocity	≥ 2.89	~ 1.51	1.06	~ 291	18.6
14	Kepler-296 e	M Warm Terran	Transit	~ 2.96	1.53	1.00	~ 281	34.1
15	Wolf 1069 b	M Warm Terran	Radial Velocity	≥ 1.26	~ 1.08	0.65	~ 258	15.6
16	TRAPPIST-1 e	M Warm Terran	Transit	0.69	0.92	0.65	~ 257	6.10
17	Kepler-442 b	K Warm Terran	Transit	~ 2.36	1.34	0.70	~ 263	112
18	Kepler-1652 b	M Warm Terran	Transit	~ 3.19	1.60	0.84	~ 275	38.1
19	K2-3 d	M Warm Terran	Transit	< 2.20	1.46	1.45	~ 315	44.6
20	TOI-715 b	M Warm Terran	Transit	~ 3.02	1.55	0.71	~ 264	19.3
21	GJ 667 C f	M Warm Terran	Radial Velocity	≥ 2.70	~ 1.45	0.56	~ 249	39.0
22	Kepler-62 f	K Warm Terran	Transit	< 35.0	1.41	0.41	~ 229	267
23	TRAPPIST-1 f	M Warm Terran	Transit	1.04	1.04	0.37	~ 224	9.21
24	Teegarden's Star c	M Warm Terran	Radial Velocity	≥ 1.05	~ 1.02	0.35	~ 221	11.4

- Habitable Worlds Catalogue

Exoplanet Habitability

Where do we get the features?

Habitable Worlds Catalogue, PHL @ UPR Arecibo



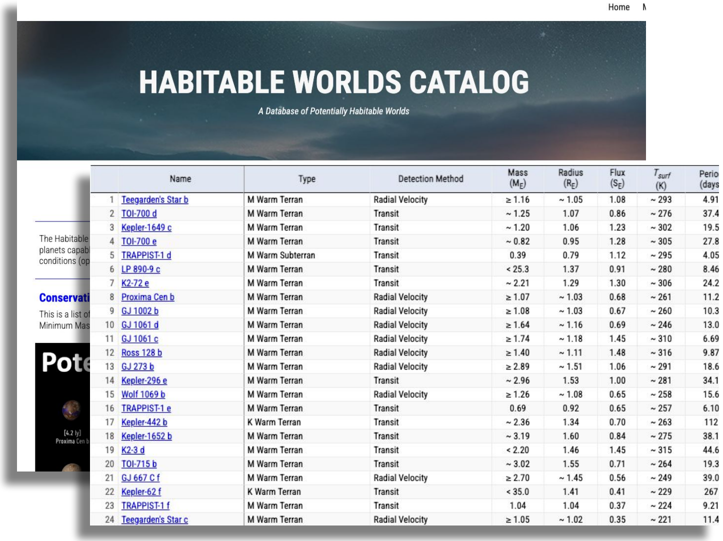
	Name	Type	Detection Method	Mass (M_E)	Radius (R_E)	Flux (S_E)	T_{surf} (K)	Perio (days
1	Teegarden's Star b	M Warm Terran	Radial Velocity	≥ 1.16	~ 1.05	1.08	~ 293	4.91
2	TOI-700 d	M Warm Terran	Transit	~ 1.25	1.07	0.86	~ 276	37.4
3	Kepler-1649 c	M Warm Terran	Transit	~ 1.20	1.06	1.23	~ 302	19.5
4	TOI-700 e	M Warm Terran	Transit	~ 0.82	0.95	1.28	~ 305	27.8
5	TRAPPIST-1 d	M Warm Subterran	Transit	0.39	0.79	1.12	~ 295	4.05
6	LP 890-9 c	M Warm Terran	Transit	< 25.3	1.37	0.91	~ 280	8.46
7	K2-72 e	M Warm Terran	Transit	~ 2.21	1.29	1.30	~ 306	24.2
8	Proxima Cen b	M Warm Terran	Radial Velocity	≥ 1.07	~ 1.03	0.68	~ 261	11.2
9	GJ 1002 b	M Warm Terran	Radial Velocity	≥ 1.08	~ 1.03	0.67	~ 260	10.3
10	GJ 1061 d	M Warm Terran	Radial Velocity	≥ 1.64	~ 1.16	0.69	~ 246	13.0
11	GJ 1061 c	M Warm Terran	Radial Velocity	≥ 1.74	~ 1.18	1.45	~ 310	6.69
12	Ross 128 b	M Warm Terran	Radial Velocity	≥ 1.40	~ 1.11	1.48	~ 316	9.87
13	GJ 273 b	M Warm Terran	Radial Velocity	≥ 2.89	~ 1.51	1.06	~ 291	18.6
14	Kepler-296 e	M Warm Terran	Transit	~ 2.96	1.53	1.00	~ 281	34.1
15	Wolf 1069 b	M Warm Terran	Radial Velocity	≥ 1.26	~ 1.08	0.65	~ 258	15.6
16	TRAPPIST-1 e	M Warm Terran	Transit	0.69	0.92	0.65	~ 257	6.10
17	Kepler-442 b	K Warm Terran	Transit	~ 2.36	1.34	0.70	~ 263	112
18	Kepler-1652 b	M Warm Terran	Transit	~ 3.19	1.60	0.84	~ 275	38.1
19	K2-3 d	M Warm Terran	Transit	< 2.20	1.46	1.45	~ 315	44.6
20	TOI-715 b	M Warm Terran	Transit	~ 3.02	1.55	0.71	~ 264	19.3
21	GJ 667 C f	M Warm Terran	Radial Velocity	≥ 2.70	~ 1.45	0.56	~ 249	39.0
22	Kepler-62 f	K Warm Terran	Transit	< 35.0	1.41	0.41	~ 229	267
23	TRAPPIST-1 f	M Warm Terran	Transit	1.04	1.04	0.37	~ 224	9.21
24	Teegarden's Star c	M Warm Terran	Radial Velocity	≥ 1.05	~ 1.02	0.35	~ 221	11.4

- Habitable Worlds Catalogue
  - 5,599 confirmed exoplanets

Exoplanet Habitability

Where do we get the features?

Habitable Worlds Catalogue, PHL @ UPR Arecibo



	Name	Type	Detection Method	Mass (M_E)	Radius (R_E)	Flux (S_E)	T_{surf} (K)	Perio (days
1	Teegarden's Star b	M Warm Terran	Radial Velocity	≥ 1.16	~ 1.05	1.08	~ 293	4.91
2	TOI-700 d	M Warm Terran	Transit	~ 1.25	1.07	0.86	~ 276	37.4
3	Kepler-1649 c	M Warm Terran	Transit	~ 1.20	1.06	1.23	~ 302	19.5
4	TOI-700 e	M Warm Terran	Transit	~ 0.82	0.95	1.28	~ 305	27.8
5	TRAPPIST-1 d	M Warm Subterran	Transit	0.39	0.79	1.12	~ 295	4.05
6	LP 890-9 c	M Warm Terran	Transit	< 25.3	1.37	0.91	~ 280	8.46
7	K2-72 e	M Warm Terran	Transit	~ 2.21	1.29	1.30	~ 306	24.2
8	Proxima Cen b	M Warm Terran	Radial Velocity	≥ 1.07	~ 1.03	0.68	~ 261	11.2
9	GJ 1002 b	M Warm Terran	Radial Velocity	≥ 1.08	~ 1.03	0.67	~ 260	10.3
10	GJ 1061 d	M Warm Terran	Radial Velocity	≥ 1.64	~ 1.16	0.69	~ 246	13.0
11	GJ 1061 c	M Warm Terran	Radial Velocity	≥ 1.74	~ 1.18	1.45	~ 310	6.69
12	Ross 128 b	M Warm Terran	Radial Velocity	≥ 1.40	~ 1.11	1.48	~ 316	9.87
13	GJ 273 b	M Warm Terran	Radial Velocity	≥ 2.89	~ 1.51	1.06	~ 291	18.6
14	Kepler-296 e	M Warm Terran	Transit	~ 2.96	1.53	1.00	~ 281	34.1
15	Wolf 1069 b	M Warm Terran	Radial Velocity	≥ 1.26	~ 1.08	0.65	~ 258	15.6
16	TRAPPIST-1 e	M Warm Terran	Transit	0.69	0.92	0.65	~ 257	6.10
17	Kepler-442 b	K Warm Terran	Transit	~ 2.36	1.34	0.70	~ 263	112
18	Kepler-1652 b	M Warm Terran	Transit	~ 3.19	1.60	0.84	~ 275	38.1
19	K2-3 d	M Warm Terran	Transit	< 2.20	1.46	1.45	~ 315	44.6
20	TOI-715 b	M Warm Terran	Transit	~ 3.02	1.55	0.71	~ 264	19.3
21	GJ 667 C f	M Warm Terran	Radial Velocity	≥ 2.70	~ 1.45	0.56	~ 249	39.0
22	Kepler-62 f	K Warm Terran	Transit	< 35.0	1.41	0.41	~ 229	267
23	TRAPPIST-1 f	M Warm Terran	Transit	1.04	1.04	0.37	~ 224	9.21
24	Teegarden's Star c	M Warm Terran	Radial Velocity	≥ 1.05	~ 1.02	0.35	~ 221	11.4

- Habitable Worlds Catalogue
  - 5,599 confirmed exoplanets
  - Includes a data field:

    [pl_habitable]

Exoplanet Habitability

Choosing the models:

https://doi.org/10.48550/arXiv.2207.08815

Why do tree-based models still outperform deep learning on tabular data?

Léo Grinsztajn
Soda, Inria Saclay
leo.grinsztajn@inria.fr

Edouard Oyallon
ISIR, CNRS, Sorbonne University

Gaël Varoquaux
Soda, Inria Saclay

Abstract

While deep learning has enabled tremendous progress on text and image datasets, its superiority on tabular data is not clear. We contribute extensive benchmarks of standard and novel deep learning methods as well as tree-based models such as XGBoost and Random Forests, across a large number of datasets and hyperparameter combinations. We define a standard set of 45 datasets from varied domains with clear characteristics of tabular data and a benchmarking methodology accounting for both fitting models and finding good hyperparameters. Results show that tree-based models remain state-of-the-art on medium-sized data (∼10K samples) even without accounting for their superior speed. To understand this gap, we conduct an empirical investigation into the differing inductive biases of tree-based models and Neural Networks (NNs). This leads to a series of challenges which should guide researchers aiming to build tabular-specific NNs: **1.** be robust to uninformative features, **2.** preserve the orientation of the data, and **3.** be able to easily learn irregular functions. To stimulate research on tabular architectures, we contribute a standard benchmark and raw data for baselines: every point of a 20 000 compute hours hyperparameter search for each learner.

1 Introduction

Deep learning has enabled tremendous progress for learning on image, language, or even audio datasets. On tabular data, however, the picture is muddier and ensemble models based on decision trees like XGBoost remain the go-to tool for most practitioners [Sta] and data science competitions [Kossen et al., 2021]. Indeed deep learning architectures have been crafted to create inductive biases matching invariances and spatial dependencies of the data. Finding corresponding invariances is hard in tabular data, made of heterogeneous features, small sample sizes, extreme values.

Creating tabular-specific deep learning architectures is a very active area of research (see section 2) given that tree-based models are not differentiable, and thus cannot be easily composed and jointly trained with other deep learning blocks. Most corresponding publications claim to beat or match tree-based models, but their claims have been put into question: a simple Resnet seems to be competitive with some of these new models [Gorishniy et al., 2021], and most of these methods seem to fail on new datasets [Shwartz-Ziv and Armon, 2021]. Indeed, the lack of an established benchmark for tabular

https://doi.org/10.48550/arXiv.2106.03253

TABULAR DATA: DEEP LEARNING IS NOT ALL YOU NEED

Ravid Shwartz-Ziv
ravid.ziv@intel.com
IT AI Group, Intel

Amitai Armon
amitai.armon@intel.com
IT AI Group, Intel

November 24, 2021

ABSTRACT

A key element in solving real-life data science problems is selecting the types of models to use. Tree ensemble models (such as XGBoost) are usually recommended for classification and regression problems with tabular data. However, several deep learning models for tabular data have recently been proposed, claiming to outperform XGBoost for some use cases. This paper explores whether these deep models should be a recommended option for tabular data by rigorously comparing the new deep models to XGBoost on various datasets. In addition to systematically comparing their performance, we consider the tuning and computation they require. Our study shows that XGBoost outperforms these deep models across the datasets, including the datasets used in the papers that proposed the deep models. We also demonstrate that XGBoost requires much less tuning. On the positive side, we show that an ensemble of deep models and XGBoost performs better on these datasets than XGBoost alone.

Keywords Tabular data · Deep neural networks · Tree-based models · Hyperparameter optimization

1 Introduction

Deep neural networks have demonstrated great success across various domains, including images, audio, and text [Devlin et al., 2019, He et al., 2016, van den Oord et al., 2016]. There are several canonical architectures for encoding raw data into meaningful representations in these domains. These canonical architectures usually perform well in real-world applications.

In real-world applications, the most common data type is tabular data, comprising samples (rows) with the same set of features (columns). Tabular data is used in practical applications in many fields, including medicine, finance, manufacturing, climate science, and many other applications that are based on relational databases. During the last decade, traditional machine learning methods, such as gradient-boosted decision trees (GBDT) [Chen and Guestrin, 2016], still dominated tabular data modeling and showed superior performance over deep learning. In spite of their theoretical advantages [Shwartz-Ziv et al., 2018, Poggio et al., 2020, Piran et al., 2020], deep neural networks pose many challenges when applied to tabular data, such as lack of locality, data sparsity (missing values), mixed feature types (numerical, ordinal, and categorical), and lack of prior knowledge about the dataset structure (unlike with text or images). Moreover, deep neural networks are perceived as a "black box" approach – in other words, they lack transparency or interpretability of how input data are transformed into model outputs [Shwartz-Ziv and Tishby, 2017]. Although the “no free lunch” principle [Wolpert and Macready, 1997] always applies, tree-ensemble algorithms, such

Exoplanet Habitability

Choosing the models:

https://doi.org/10.48550/arXiv.2207.08815

Why do tree-based models still outperform deep learning on tabular data?

Léo Grinsztajn
Soda, Inria Saclay
leo.grinsztajn@inria.fr

Edouard Oyallon
ISIR, CNRS, Sorbonne University

Gaël Varoquaux
Soda, Inria Saclay

Abstract

While deep learning has enabled tremendous progress on text and image datasets, its superiority on tabular data is not clear. We contribute extensive benchmarks of standard and novel deep learning methods as well as tree-based models such as XGBoost and Random Forests, across a large number of datasets and hyperparameter combinations. We define a standard set of 45 datasets from varied domains with clear characteristics of tabular data and a benchmarking methodology accounting for both fitting models and finding good hyperparameters. Results show that tree-based models remain state-of-the-art on medium-sized data (~10K samples) even without accounting for their superior speed. To understand this gap, we conduct an empirical investigation into the differing inductive biases of tree-based models and Neural Networks (NNs). This leads to a series of challenges which should guide researchers aiming to build tabular-specific NNs: **1.** be robust to uninformative features, **2.** preserve the orientation of the data, and **3.** be able to easily learn irregular functions. To stimulate research on tabular architectures, we contribute a standard benchmark and raw data for baselines: every point of a 20 000 compute hours hyperparameter search for each learner.

1 Introduction

Deep learning has enabled tremendous progress for learning on image, language, or even audio datasets. On tabular data, however, the picture is muddier and ensemble models based on decision trees like XGBoost remain the go-to tool for most practitioners [Sta] and data science competitions [Kossen et al., 2021]. Indeed deep learning architectures have been crafted to create inductive biases matching invariances and spatial dependencies of the data. Finding corresponding invariances is hard in tabular data, made of heterogeneous features, small sample sizes, extreme values.

Creating tabular-specific deep learning architectures is a very active area of research (see section 2) given that tree-based models are not differentiable, and thus cannot be easily composed and jointly trained with other deep learning blocks. Most corresponding publications claim to beat or match tree-based models, but their claims have been put into question: a simple Resnet seems to be competitive with some of these new models [Gorishniy et al., 2021], and most of these methods seem to fail on new datasets [Shwartz-Ziv and Armon, 2021]. Indeed, the lack of an established benchmark for tabular

Random Forest

XGBoost

https://doi.org/10.48550/arXiv.2106.03253

TABULAR DATA: DEEP LEARNING IS NOT ALL YOU NEED

Ravid Shwartz-Ziv
ravid.ziv@intel.com
IT AI Group, Intel

Amitai Armon
amitai.armon@intel.com
IT AI Group, Intel

November 24, 2021

ABSTRACT

A key element in solving real-life data science problems is selecting the types of models to use. Tree ensemble models (such as XGBoost) are usually recommended for classification and regression problems with tabular data. However, several deep learning models for tabular data have recently been proposed, claiming to outperform XGBoost for some use cases. This paper explores whether these deep models should be a recommended option for tabular data by rigorously comparing the new deep models to XGBoost on various datasets. In addition to systematically comparing their performance, we consider the tuning and computation they require. Our study shows that XGBoost outperforms these deep models across the datasets, including the datasets used in the papers that proposed the deep models. We also demonstrate that XGBoost requires much less tuning. On the positive side, we show that an ensemble of deep models and XGBoost performs better on these datasets than XGBoost alone.

Keywords Tabular data · Deep neural networks · Tree-based models · Hyperparameter optimization

1 Introduction

Deep neural networks have demonstrated great success across various domains, including images, audio, and text [Devlin et al., 2019, He et al., 2016, van den Oord et al., 2016]. There are several canonical architectures for encoding raw data into meaningful representations in these domains. These canonical architectures usually perform well in real-world applications.

In real-world applications, the most common data type is tabular data, comprising samples (rows) with the same set of features (columns). Tabular data is used in practical applications in many fields, including medicine, finance, manufacturing, climate science, and many other applications that are based on relational databases. During the last decade, traditional machine learning methods, such as gradient-boosted decision trees (GBDT) [Chen and Guestrin, 2016], still dominated tabular data modeling and showed superior performance over deep learning. In spite of their theoretical advantages [Shwartz-Ziv et al., 2018, Poggio et al., 2020, Piran et al., 2020], deep neural networks pose many challenges when applied to tabular data, such as lack of locality, data sparsity (missing values), mixed feature types (numerical, ordinal, and categorical), and lack of prior knowledge about the dataset structure (unlike with text or images). Moreover, deep neural networks are perceived as a "black box" approach – in other words, they lack transparency or interpretability of how input data are transformed into model outputs [Shwartz-Ziv and Tishby, 2017]. Although the "no free lunch" principle [Wolpert and Macready, 1997] always applies, tree-ensemble algorithms, such

Exoplanet Habitability

Step 1: Data Downloading

- ***NASA Exoplanet Archive*** *(**5,834** exoplanets)*
- ***Habitable Worlds Catalog (HWC)****, PHL @ UPR Arecibo (**5,599** exoplanets)*

Join data from NASA and HWC.

Use **P-HABITABLE** data field from HWC to label training data.

Step 2: Feature Engineering

Data Joining

Data Cleaning

* Filter out exoplanets with questionable status. (**pl_controv_flag**=1)

* Limit dataset to exoplanets. with **single-host stars**.

* Remove irrelevant features.

Handling Missing Values

* Remove features with ≥25% missing values.

* **Categorical**: fill missing values with **mode**.

* **Numerical**: fill missing values with Multivariate Imputation by Chained Equation (**MICE**).

Converting Categorical Values

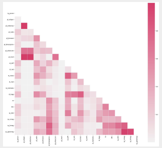
* **One-Hot Encoding** to convert categorical features into numerical values.

Handling Data Imbalance

* The dataset is highly imbalanced. Apply a combination of Synthetic Minority OVersampling Technique (**SMOTE**) and Edited Nearest Neighbors (**ENN**) techniques to oversample minorities and downsample majorities.

Data Correlation Analysis

Remove redundant features based on **data correlation analysis**.



Min-Max Scalar

Standardize value ranges to **[0.1]**

Step 3: Model Training

* Randomly shuffle dataset, then split into train & test datasets for training & evaluation. Repeat 100 times. Use avg accuracy as the measure of model quality for the corresponding hyperparameters.

* Hyperparameter tuning through exhaustive search.

Step 4: SHAP Analysis

* Feature importance analysis through **SHAP**.

* Perform analysis through **SHAP** to understand how the influential features impact the exoplanet habitability.

Random Forest Model Evaluation

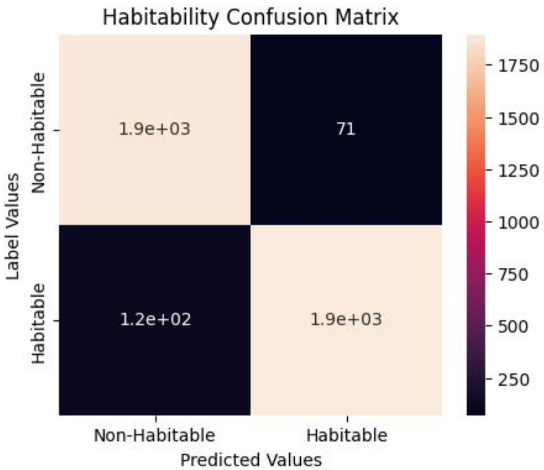
Class	Random Forest		
	Precision	Recall	F1 score
1 (Habitable)	0.96	0.94	0.95
0 (Non-Habitable)	0.94	0.96	0.95

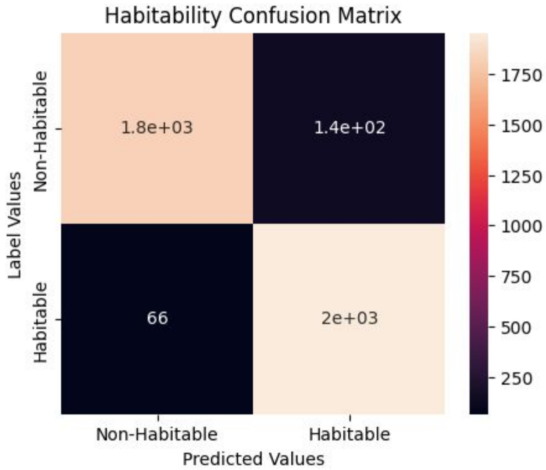
XGBoost Model Evaluation

Class	XGBoost		
	Precision	Recall	F1 score
1 (Habitable)	0.94	0.97	0.95
0 (Non-Habitable)	0.97	0.93	0.95

Exoplanet Habitability

Model results:





Random Forest Classifier - Classification Report:

	precision	recall	f1-score	support
0.0	0.94	0.96	0.95	1959
1.0	0.96	0.94	0.95	2020
accuracy			0.95	3979

XGBoost Classifier - Classification Report:

	precision	recall	f1-score	support
0.0	0.97	0.93	0.95	1959
1.0	0.94	0.97	0.95	2020
accuracy			0.95	3979

Exoplanet **Habitability**

Exoplanet **Habitability**

Introducing **SHAP.**

Exoplanet Habitability

Introducing **SHAP.**

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

Exoplanet Habitability

Introducing **SHAP.**

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

✦ How it basically works:

Exoplanet Habitability

Introducing **SHAP.**

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

- How it basically works:
  - Look at each feature (ex. feature 1) of the dataset and train the model once with the feature and once without the feature, for each subset of the rest of the features in your training set.

Exoplanet Habitability

Introducing **SHAP.**

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

- ✦ How it basically works:
  - ✧ Look at each feature (ex. feature 1) of the dataset and train the model once with the feature and once without the feature, for each subset of the rest of the features in your training set.
  - ✧ Each of these evaluations are applied to test dataset. Then, find the models' predicted difference.

Exoplanet Habitability

Introducing **SHAP.**

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

- ✦ How it basically works:
  - ✧ Look at each feature (ex. feature 1) of the dataset and train the model once with the feature and once without the feature, for each subset of the rest of the features in your training set.
  - ✧ Each of these evaluations are applied to test dataset. Then, find the models' predicted difference.
  - ✧ Find weighted average across the diff. subsets to get **feature 1's SHAP value.**

Exoplanet Habitability

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

- ✦ ML models sometimes like a black box. SHAP helps you decode it.

Exoplanet Habitability

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

- ✦ ML models sometimes like a black box. SHAP helps you decode it.
- ✦ Apply to any model! (Because of the nature of SHAP)

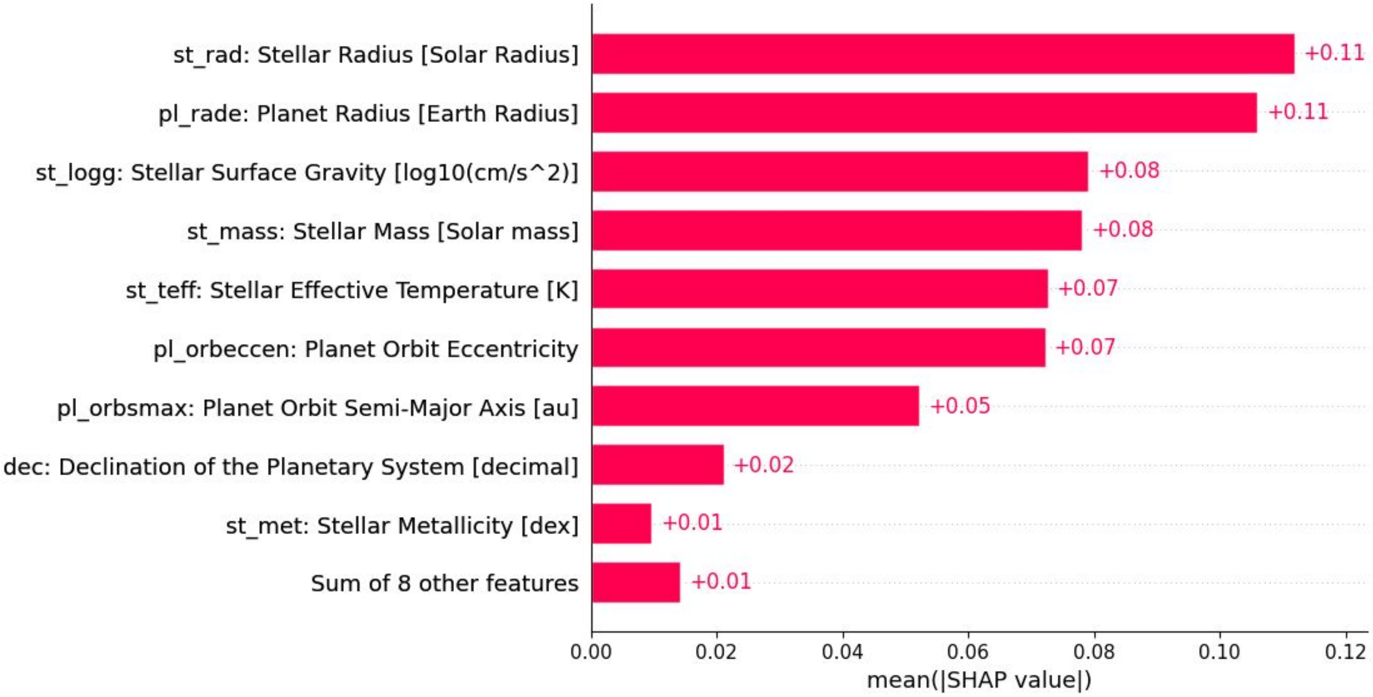
Exoplanet Habitability

SHAP (Shapley Additive Explanations)
✦ A framework that helps us to recognize the impact of each individual feature positively or negatively affecting our outcome.

- ✦ ML models sometimes like a black box. SHAP helps you decode it.
- ✦ Apply to any model! (Because of the nature of SHAP)
- ✦ Also, **ADDITIVE.**

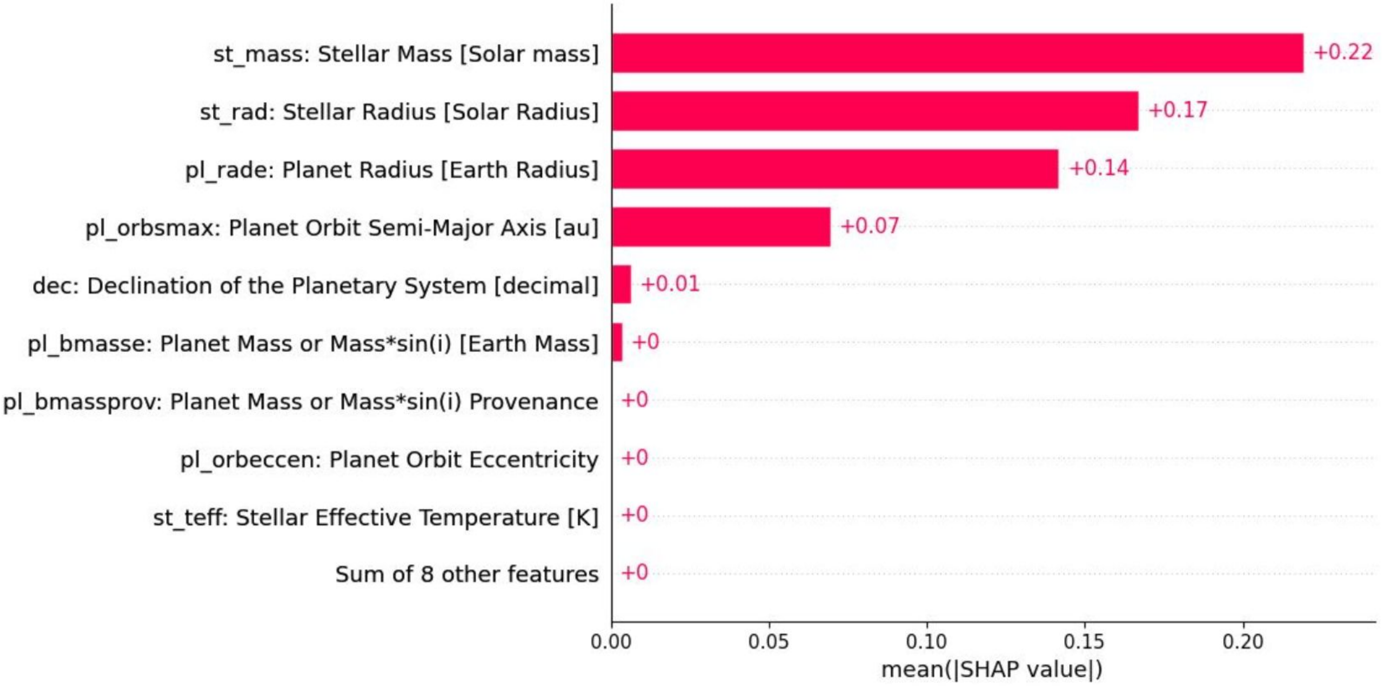
Exoplanet Habitability

Overall Analysis – **Random Forest**



Exoplanet Habitability

Overall Analysis – **XGBoost**



Exoplanet Habitability

You can go further with SHAP!!

Exoplanet Habitability

You can go further with SHAP!!

- ✦ Can analyze whether a feature pushes the exoplanet **towards** habitability or **away**.

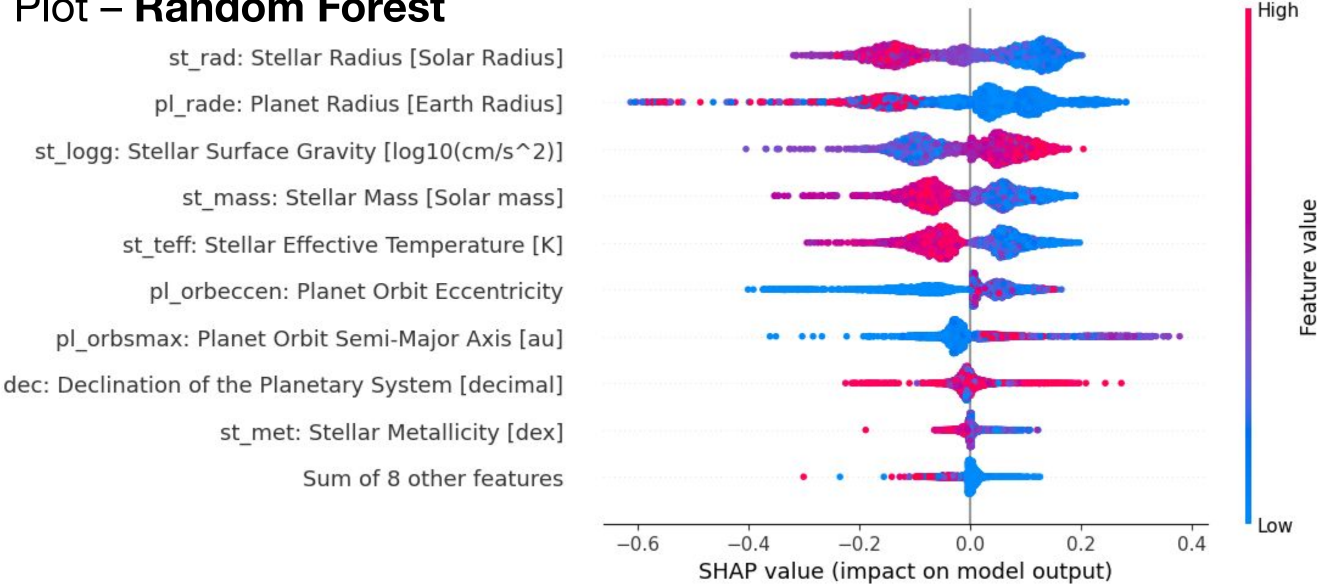
Exoplanet Habitability

You can go further with SHAP!!

- Can analyze whether a feature pushes the exoplanet **towards** habitability or **away** from it.
  - Positive SHAP value → towards.
  - Negative SHAP value → away.

Exoplanet Habitability

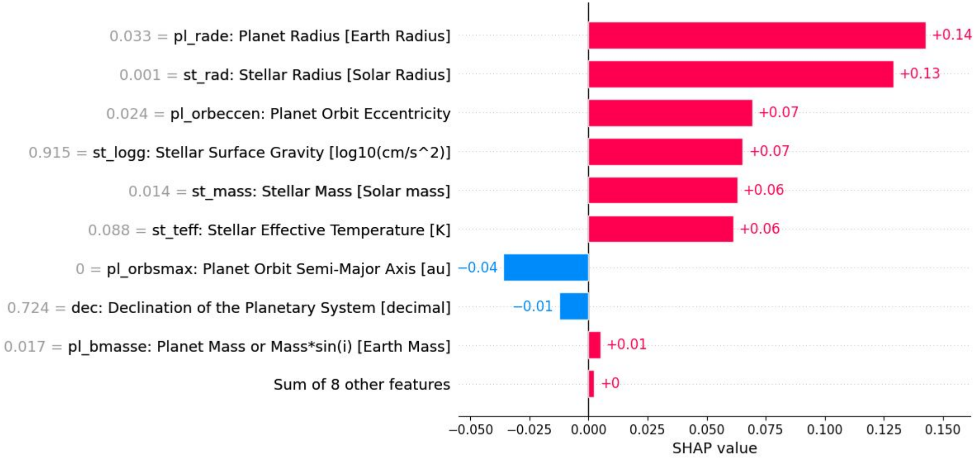
Beeswarm Plot – **Random Forest**



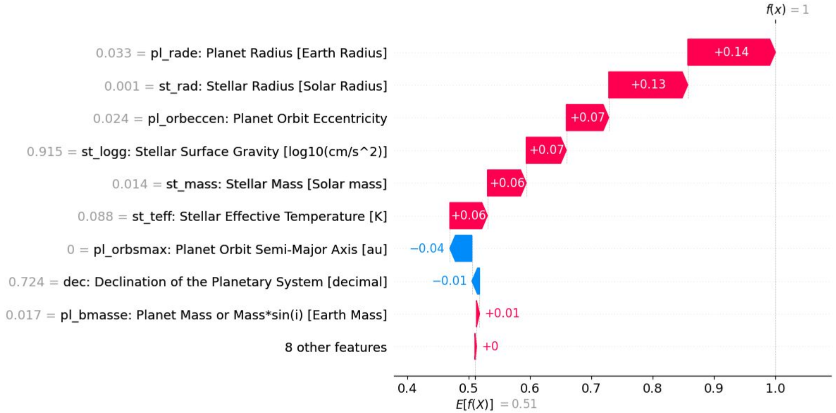
- Higher values (relative to other samples in dataset) of stellar radius, planet radius, stellar mass, and stellar effective temperature lead towards negative predictions, while lower values lead towards positive outcomes.
- Planet orbit semi-major axis, on the other hand, has the opposite impact on prediction outcomes, with higher values leading toward positive predictions while lower values leading towards negative outcomes.

Exoplanet Habitability

Local Plots – **Random Forest**



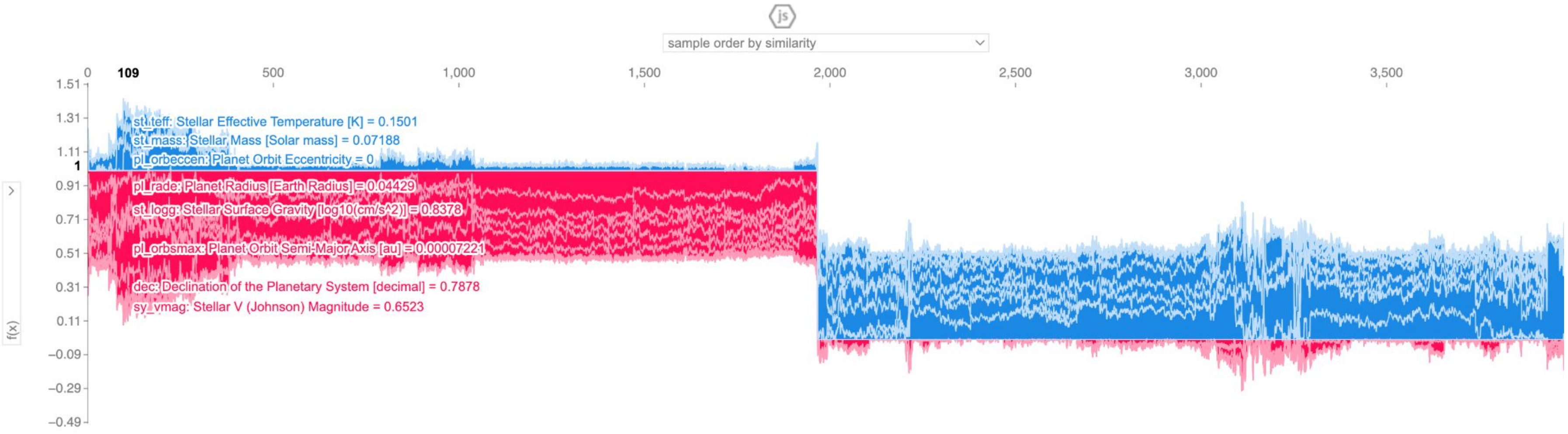
SHAP local bar plot for one sample in the dataset



SHAP waterfall plot for one sample in the dataset

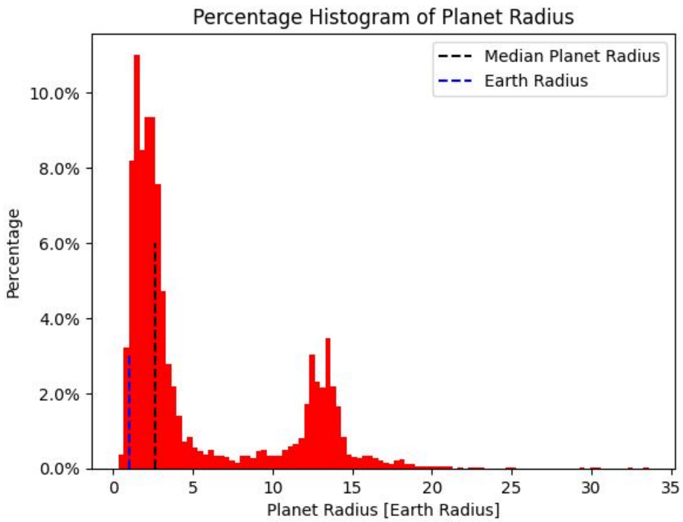
Exoplanet Habitability

Force Plots – **Random Forest**

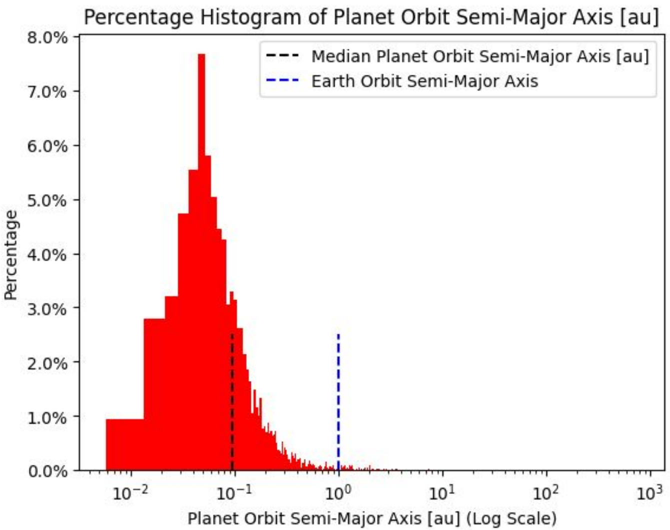


Exoplanet Habitability

Sanity check: does all this align with reality?



SHAP analysis indicates a higher planet radius leads towards negative predictions, while lower value leads towards positive predictions.

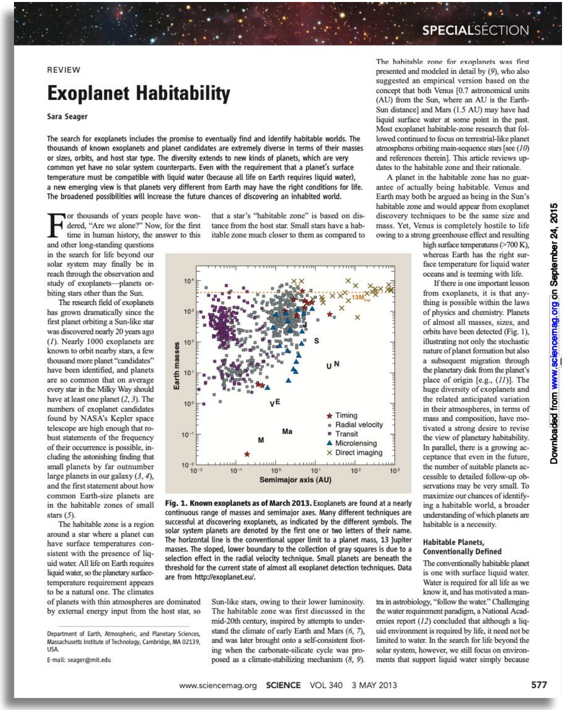


SHAP analysis indicates a higher planet orbit semi-major axis leads towards positive predictions, while lower value leads towards negative predictions.

Exoplanet **Classification**

New Paper

Prof. Seager’s 2013 Paper



SPECIALSECTION

REVIEW

Exoplanet Habitability

Sara Seager

The search for exoplanets includes the promise to eventually find and identify habitable worlds. The thousands of known exoplanets and planet candidates are extremely diverse in terms of their masses or sizes, orbits, and host star type. The diversity extends to new kinds of planets, which are very common yet have no solar system counterparts. Even with the requirement that a planet’s surface temperature must be compatible with liquid water (because all life on Earth requires liquid water), a new emerging view is that planets very different from Earth may have the right conditions for life. The broadened possibilities will increase the future chances of discovering an inhabited world.

For thousands of years people have wondered, “Are we alone?” Now, for the first time in human history, the answer to this and other long-standing questions in the search for life beyond our solar system may finally be in reach through the observation and study of exoplanets—planets orbiting stars other than the Sun.

The research field of exoplanets has grown dramatically since the first planet orbiting a Sun-like star was discovered nearly 20 years ago (*1*). Nearly 1000 exoplanets are known to orbit nearby stars, a few thousand more planet “candidates” have been identified, and planets are so common that on average every star in the Milky Way should have at least one planet (*2*, *3*). The numbers of exoplanet candidates found by NASA’s Kepler space telescope are high enough that robust statements of the frequency of their occurrence is possible, including the astonishing finding that small planets by far outnumber large planets in our galaxy (*3*, *4*), and the first statement about how common Earth-size planets are in the habitable zones of small stars (*5*).

The habitable zone is a region around a star where a planet can have surface temperatures consistent with the presence of liquid water. All life on Earth requires liquid water, so the planetary surface-temperature requirement appears to be a natural one. The climates of planets with thin atmospheres are dominated by external energy input from the host star, so that a star’s “habitable zone” is based on distance from the host star. Small stars have a habitable zone much closer to them as compared to Sun-like stars, owing to their lower luminosity. The habitable zone was first discussed in the mid-20th century, inspired by attempts to understand the climate of early Earth and Mars (*6*, *7*), and was later brought onto a self-consistent footing when the carbonate-silicate cycle was proposed as a climate-stabilizing mechanism (*8*, *9*).

The habitable zone for exoplanets was first presented and modeled in detail by (*9*), who also suggested an empirical version based on the concept that both Venus [0.7 astronomical units (AU) from the Sun, where an AU is the Earth-Sun distance] and Mars (1.5 AU) may have had liquid surface water at some point in the past. Most exoplanet habitable-zone research that followed continued to focus on terrestrial-like planet atmospheres orbiting main-sequence stars [see (*10*) and references therein]. This article reviews updates to the habitable zone and their rationale.

A planet in the habitable zone has no guarantee of actually being habitable. Venus and Earth may both be argued as being in the Sun’s habitable zone and would appear from exoplanet discovery techniques to be the same size and mass. Yet, Venus is completely hostile to life owing to a strong greenhouse effect and resulting high surface temperatures (>700 K), whereas Earth has the right surface temperature for liquid water oceans and is teeming with life.

If there is one important lesson from exoplanets, it is that anything is possible within the laws of physics and chemistry. Planets of almost all masses, sizes, and orbits have been detected (Fig. 1), illustrating not only the stochastic nature of planet formation but also a subsequent migration through the planetary disk from the planet’s place of origin [e.g., (*11*)]. The huge diversity of exoplanets and the related anticipated variation in their atmospheres, in terms of mass and composition, have motivated a strong desire to revise the view of planetary habitability. In parallel, there is a growing acceptance that even in the future, the number of suitable planets accessible to detailed follow-up observations may be very small. To maximize our chances of identifying a habitable world, a broader understanding of which planets are habitable is a necessity.

Habitable Planets, Conventionally Defined

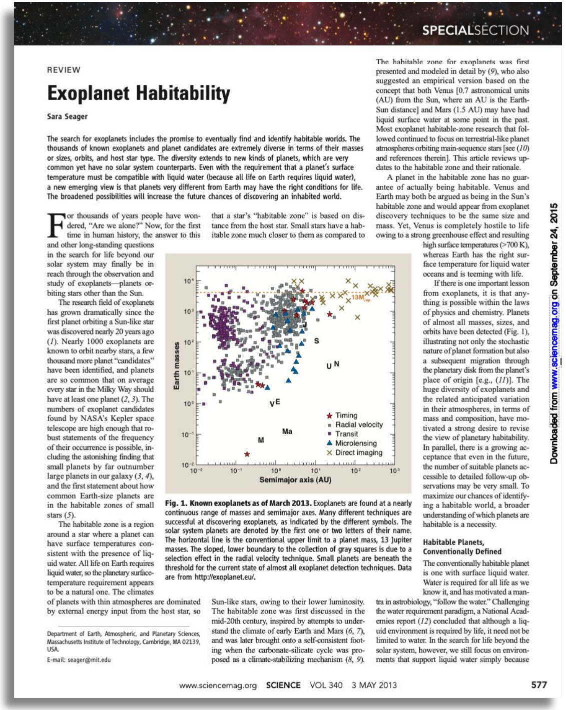
The conventionally habitable planet is one with surface liquid water. Water is required for all life as we know it, and has motivated a mantra in astrobiology, “follow the water.” Challenging the water requirement paradigm, a National Academies report (*12*) concluded that although a liquid environment is required by life, it need not be limited to water. In the search for life beyond the solar system, however, we still focus on environments that support liquid water simply because

Fig. 1. Known exoplanets as of March 2013. Exoplanets are found at a nearly continuous range of masses and semimajor axes. Many different techniques are successful at discovering exoplanets, as indicated by the different symbols. The solar system planets are denoted by the first one or two letters of their name. The horizontal line is the conventional upper limit to a planet mass, 13 Jupiter masses. The sloped, lower boundary to the collection of gray squares is due to a selection effect in the radial velocity technique. Small planets are beneath the threshold for the current state of almost all exoplanet detection techniques. Data are from http://exoplanet.eu/.

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E-mail: seager@mit.edu

New Paper

Prof. Seager's 2013 Paper



REVIEW

Exoplanet Habitability

Sara Seager

The search for exoplanets includes the promise to eventually find and identify habitable worlds. The thousands of known exoplanets and planet candidates are extremely diverse in terms of their masses or sizes, orbits, and host star type. The diversity extends to new kinds of planets, which are very common yet have no solar system counterparts. Even with the requirement that a planet's surface temperature must be compatible with liquid water (because all life on Earth requires liquid water), a new emerging view is that planets very different from Earth may have the right conditions for life. The broadened possibilities will increase the future chances of discovering an inhabited world.

For thousands of years people have wondered, "Are we alone?" Now, for the first time in human history, the answer to this and other long-standing questions in the search for life beyond our solar system may finally be in reach through the observation and study of exoplanets—planets orbiting stars other than the Sun.

The research field of exoplanets has grown dramatically since the first planet orbiting a Sun-like star was discovered nearly 20 years ago (*1*). Nearly 1000 exoplanets are known to orbit nearby stars, a few thousand more planet "candidates" have been identified, and planets are so common that on average every star in the Milky Way should have at least one planet (*2*, *3*). The numbers of exoplanet candidates found by NASA's Kepler space telescope are high enough that robust statements of the frequency of their occurrence is possible, including the astonishing finding that small planets by far outnumber large planets in our galaxy (*3*, *4*), and the first statement about how common Earth-size planets are in the habitable zones of small stars (*5*).

The habitable zone is a region around a star where a planet can have surface temperatures consistent with the presence of liquid water. All life on Earth requires liquid water, so the planetary surface-temperature requirement appears to be a natural one. The climates of planets with thin atmospheres are dominated by external energy input from the host star, so that a star's "habitable zone" is based on distance from the host star. Small stars have a habitable zone much closer to them as compared to Sun-like stars, owing to their lower luminosity. The habitable zone was first discussed in the mid-20th century, inspired by attempts to understand the climate of early Earth and Mars (*6*, *7*), and was later brought onto a self-consistent footing when the carbonate-silicate cycle was proposed as a climate-stabilizing mechanism (*8*, *9*).

The habitable zone for exoplanets was first presented and modeled in detail by (*9*), who also suggested an empirical version based on the concept that both Venus [0.7 astronomical units (AU) from the Sun, where an AU is the Earth-Sun distance] and Mars (1.5 AU) may have had liquid surface water at some point in the past. Most exoplanet habitable-zone research that followed continued to focus on terrestrial-like planet atmospheres orbiting main-sequence stars [see (*10*) and references therein]. This article reviews updates to the habitable zone and their rationale.

A planet in the habitable zone has no guarantee of actually being habitable. Venus and Earth may both be argued as being in the Sun's habitable zone and would appear from exoplanet discovery techniques to be the same size and mass. Yet, Venus is completely hostile to life owing to a strong greenhouse effect and resulting high surface temperatures (>700 K), whereas Earth has the right surface temperature for liquid water oceans and is teeming with life.

If there is one important lesson from exoplanets, it is that anything is possible within the laws of physics and chemistry. Planets of almost all masses, sizes, and orbits have been detected (Fig. 1), illustrating not only the stochastic nature of planet formation but also a subsequent migration through the planetary disk from the planet's place of origin [e.g., (*11*)]. The huge diversity of exoplanets and the related anticipated variation in their atmospheres, in terms of mass and composition, have motivated a strong desire to revise the view of planetary habitability. In parallel, there is a growing acceptance that even in the future, the number of suitable planets accessible to detailed follow-up observations may be very small. To maximize our chances of identifying a habitable world, a broader understanding of which planets are habitable is a necessity.

Habitable Planets, Conventionally Defined

The conventionally habitable planet is one with surface liquid water. Water is required for all life as we know it, and has motivated a mantra in astrobiology, "follow the water." Challenging the water requirement paradigm, a National Academies report (*12*) concluded that although a liquid environment is required by life, it need not be limited to water. In the search for life beyond the solar system, however, we still focus on environments that support liquid water simply because

Fig. 1. Known exoplanets as of March 2013. Exoplanets are found at a nearly continuous range of masses and semimajor axes. Many different techniques are successful at discovering exoplanets, as indicated by the different symbols. The solar system planets are denoted by the first one or two letters of their name. The horizontal line is the conventional upper limit to a planet mass, 13 Jupiter masses. The sloped, lower boundary to the collection of gray squares is due to a selection effect in the radial velocity technique. Small planets are beneath the threshold for the current state of almost all exoplanet detection techniques. Data are from http://exoplanet.eu/.

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E-mail: seager@mit.edu

✦ Emphasizes wide diversity of exoplanets (mass, size, orbital configurations)

New Paper

Prof. Seager's 2013 Paper

SPECIALSECTION

REVIEW

Exoplanet Habitability

Sara Seager

The search for exoplanets includes the promise to eventually find and identify habitable worlds. The thousands of known exoplanets and planet candidates are extremely diverse in terms of their masses or sizes, orbits, and host star type. The diversity extends to new kinds of planets, which are very common yet have no solar system counterparts. Even with the requirement that a planet's surface temperature must be compatible with liquid water (because all life on Earth requires liquid water), a new emerging view is that planets very different from Earth may have the right conditions for life. The broadened possibilities will increase the future chances of discovering an inhabited world.

For thousands of years people have wondered, "Are we alone?" Now, for the first time in human history, the answer to this and other long-standing questions in the search for life beyond our solar system may finally be in reach through the observation and study of exoplanets—planets orbiting stars other than the Sun.

The research field of exoplanets has grown dramatically since the first planet orbiting a Sun-like star was discovered nearly 20 years ago (1). Nearly 1000 exoplanets are known to orbit nearby stars, a few thousand more planet "candidates" have been identified, and planets are so common that on average every star in the Milky Way should have at least one planet (2, 3). The numbers of exoplanet candidates found by NASA's Kepler space telescope are high enough that robust statements of the frequency of their occurrence is possible, including the astonishing finding that small planets by far outnumber large planets in our galaxy (3, 4), and the first statement about how common Earth-size planets are in the habitable zones of small stars (5).

The habitable zone is a region around a star where a planet can have surface temperatures consistent with the presence of liquid water. All life on Earth requires liquid water, so the planetary surface-temperature requirement appears to be a natural one. The climates of planets with thin atmospheres are dominated by external energy input from the host star, so that a star's "habitable zone" is based on distance from the host star. Small stars have a habitable zone much closer to them as compared to Sun-like stars, owing to their lower luminosity. The habitable zone was first discussed in the mid-20th century, inspired by attempts to understand the climate of early Earth and Mars (6, 7), and was later brought onto a self-consistent footing when the carbonate-silicate cycle was proposed as a climate-stabilizing mechanism (8, 9).

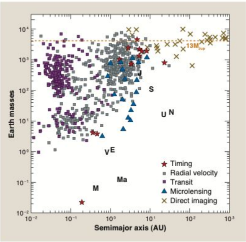


Fig. 1. Known exoplanets as of March 2013. Exoplanets are found at a nearly continuous range of masses and semimajor axes. Many different techniques are successful at discovering exoplanets, as indicated by the different symbols. The solar system planets are denoted by the first one or two letters of their name. The horizontal line is the conventional upper limit to a planet mass, 13 Jupiter masses. The sloped, lower boundary to the collection of gray squares is due to a selection effect in the radial velocity technique. Small planets are beneath the threshold for the current state of almost all exoplanet detection techniques. Data are from http://exoplanet.eu/.

The habitable zone for exoplanets was first presented and modeled in detail by (9), who also suggested an empirical version based on the concept that both Venus [0.7 astronomical units (AU) from the Sun, where an AU is the Earth-Sun distance] and Mars (1.5 AU) may have had liquid surface water at some point in the past. Most exoplanet habitable-zone research that followed continued to focus on terrestrial-like planet atmospheres orbiting main-sequence stars [see (10) and references therein]. This article reviews updates to the habitable zone and their rationale.

A planet in the habitable zone has no guarantee of actually being habitable. Venus and Earth may both be argued as being in the Sun's habitable zone and would appear from exoplanet discovery techniques to be the same size and mass. Yet, Venus is completely hostile to life owing to a strong greenhouse effect and resulting high surface temperatures (>700 K), whereas Earth has the right surface temperature for liquid water oceans and is teeming with life.

If there is one important lesson from exoplanets, it is that anything is possible within the laws of physics and chemistry. Planets of almost all masses, sizes, and orbits have been detected (Fig. 1), illustrating not only the stochastic nature of planet formation but also a subsequent migration through the planetary disk from the planet's place of origin [e.g., (11)]. The huge diversity of exoplanets and the related anticipated variation in their atmospheres, in terms of mass and composition, have motivated a strong desire to revise the view of planetary habitability. In parallel, there is a growing acceptance that even in the future, the number of suitable planets accessible to detailed follow-up observations may be very small. To maximize our chances of identifying a habitable world, a broader understanding of which planets are habitable is a necessity.

Habitable Planets, Conventionally Defined

The conventionally habitable planet is one with surface liquid water. Water is required for all life as we know it, and has motivated a mantra in astrobiology, "follow the water." Challenging the water requirement paradigm, a National Academies report (12) concluded that although a liquid environment is required by life, it need not be limited to water. In the search for life beyond the solar system, however, we still focus on environments that support liquid water simply because

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E-mail: seager@mit.edu

- ✦ Emphasizes wide diversity of exoplanets (mass, size, orbital configurations)
- ✦ Discusses difficulties in defining habitability

New Paper

Prof. Seager's 2013 Paper

SPECIALSECTION

REVIEW

Exoplanet Habitability

Sara Seager

The search for exoplanets includes the promise to eventually find and identify habitable worlds. The thousands of known exoplanets and planet candidates are extremely diverse in terms of their masses or sizes, orbits, and host star type. The diversity extends to new kinds of planets, which are very common yet have no solar system counterparts. Even with the requirement that a planet's surface temperature must be compatible with liquid water (because all life on Earth requires liquid water), a new emerging view is that planets very different from Earth may have the right conditions for life. The broadened possibilities will increase the future chances of discovering an inhabited world.

For thousands of years people have wondered, "Are we alone?" Now, for the first time in human history, the answer to this and other long-standing questions in the search for life beyond our solar system may finally be in reach through the observation and study of exoplanets—planets orbiting stars other than the Sun.

The research field of exoplanets has grown dramatically since the first planet orbiting a Sun-like star was discovered nearly 20 years ago (*1*). Nearly 1000 exoplanets are known to orbit nearby stars, a few thousand more planet "candidates" have been identified, and planets are so common that on average every star in the Milky Way should have at least one planet (*2*, *3*). The numbers of exoplanet candidates found by NASA's Kepler space telescope are high enough that robust statements of the frequency of their occurrence is possible, including the astonishing finding that small planets by far outnumber large planets in our galaxy (*3*, *4*), and the first statement about how common Earth-size planets are in the habitable zones of small stars (*5*).

The habitable zone is a region around a star where a planet can have surface temperatures consistent with the presence of liquid water. All life on Earth requires liquid water, so the planetary surface-temperature requirement appears to be a natural one. The climates of planets with thin atmospheres are dominated by external energy input from the host star, so that a star's "habitable zone" is based on distance from the host star. Small stars have a habitable zone much closer to them as compared to Sun-like stars, owing to their lower luminosity. The habitable zone was first discussed in the mid-20th century, inspired by attempts to understand the climate of early Earth and Mars (*6*, *7*), and was later brought onto a self-consistent footing when the carbonate-silicate cycle was proposed as a climate-stabilizing mechanism (*8*, *9*). The habitable zone for exoplanets was first presented and modeled in detail by (*9*), who also suggested an empirical version based on the concept that both Venus [0.7 astronomical units (AU) from the Sun, where an AU is the Earth-Sun distance] and Mars (1.5 AU) may have had liquid surface water at some point in the past. Most exoplanet habitable-zone research that followed continued to focus on terrestrial-like planet atmospheres orbiting main-sequence stars [see (*10*) and references therein]. This article reviews updates to the habitable zone and their rationale.

A planet in the habitable zone has no guarantee of actually being habitable. Venus and Earth may both be argued as being in the Sun's habitable zone and would appear from exoplanet discovery techniques to be the same size and mass. Yet, Venus is completely hostile to life owing to a strong greenhouse effect and resulting high surface temperatures (>700 K), whereas Earth has the right surface temperature for liquid water oceans and is teeming with life.

If there is one important lesson from exoplanets, it is that anything is possible within the laws of physics and chemistry. Planets of almost all masses, sizes, and orbits have been detected (Fig. 1), illustrating not only the stochastic nature of planet formation but also a subsequent migration through the planetary disk from the planet's place of origin [e.g., (*11*)]. The huge diversity of exoplanets and the related anticipated variation in their atmospheres, in terms of mass and composition, have motivated a strong desire to revise the view of planetary habitability. In parallel, there is a growing acceptance that even in the future, the number of suitable planets accessible to detailed follow-up observations may be very small. To maximize our chances of identifying a habitable world, a broader understanding of which planets are habitable is a necessity.

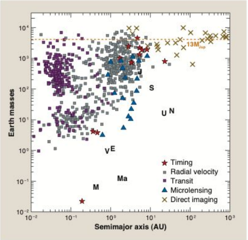


Fig. 1. Known exoplanets as of March 2013. Exoplanets are found at a nearly continuous range of masses and semimajor axes. Many different techniques are successful at discovering exoplanets, as indicated by the different symbols. The solar system planets are denoted by the first one or two letters of their name. The horizontal line is the conventional upper limit to a planet mass, 13 Jupiter masses. The sloped, lower boundary to the collection of gray squares is due to a selection effect in the radial velocity technique. Small planets are beneath the threshold for the current state of almost all exoplanet detection techniques. Data are from http://exoplanet.eu/.

Habitable Planets, Conventionally Defined

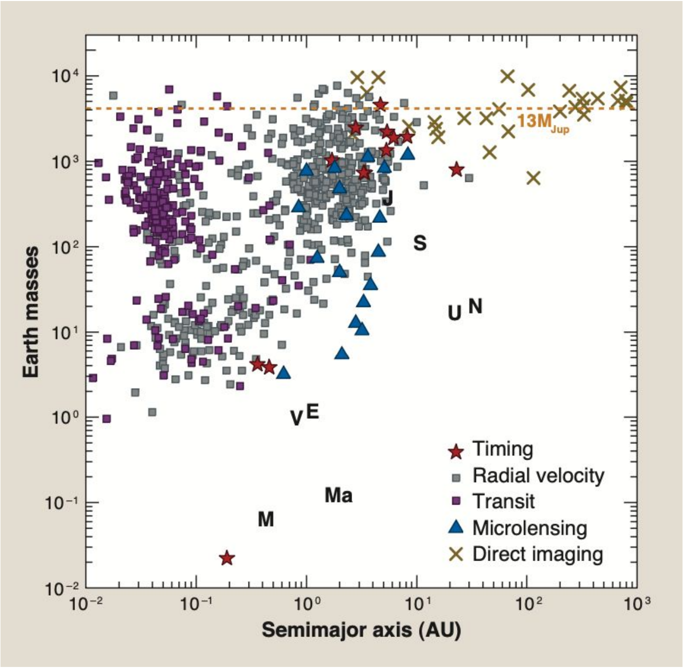
The conventionally habitable planet is one with surface liquid water. Water is required for all life as we know it, and has motivated a mantra in astrobiology, "follow the water." Challenging the water requirement paradigm, a National Academies report (*12*) concluded that although a liquid environment is required by life, it need not be limited to water. In the search for life beyond the solar system, however, we still focus on environments that support liquid water simply because

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E-mail: seager@mit.edu

- ✦ Emphasizes wide diversity of exoplanets (mass, size, orbital configurations)
- ✦ Discusses difficulties in defining habitability
- ✦ At time of publication, **only 1000 exoplanets had been discovered.**

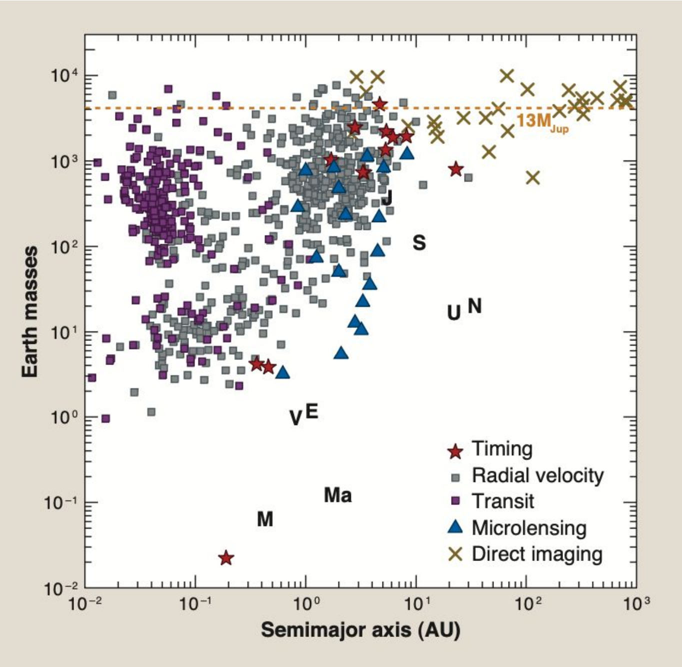
New Paper

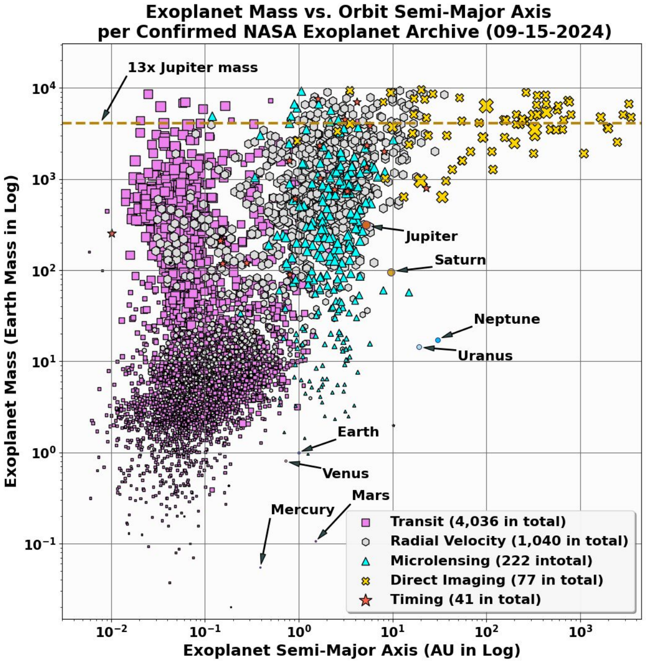
Prof. Seager's 2013 Paper



New Paper

Prof. Seager's 2013 Paper

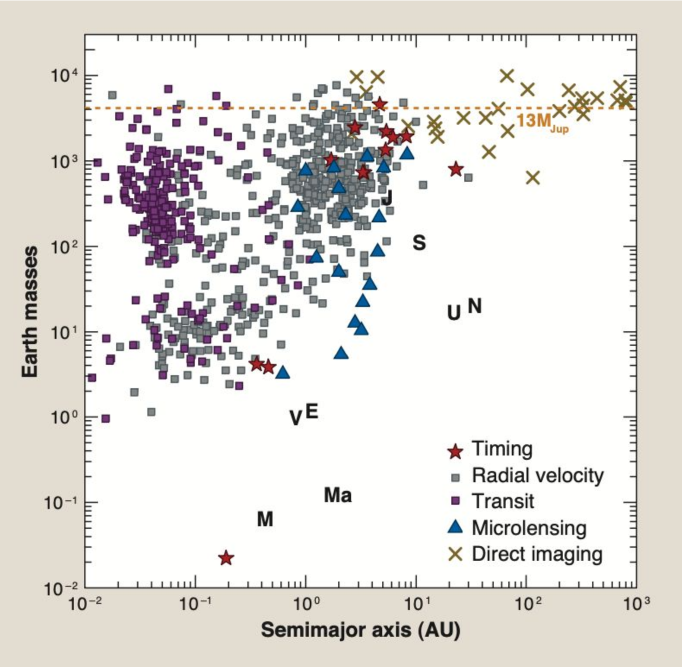


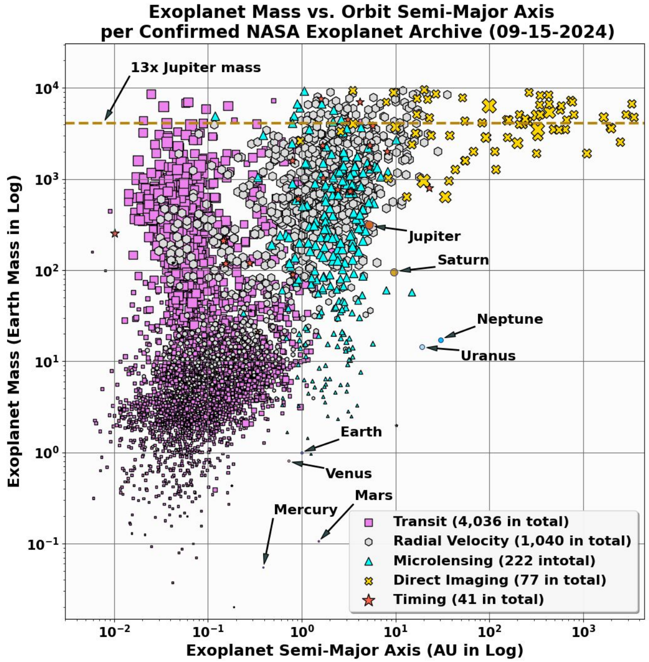


✦ Included solar system figures

New Paper

Prof. Seager’s 2013 Paper

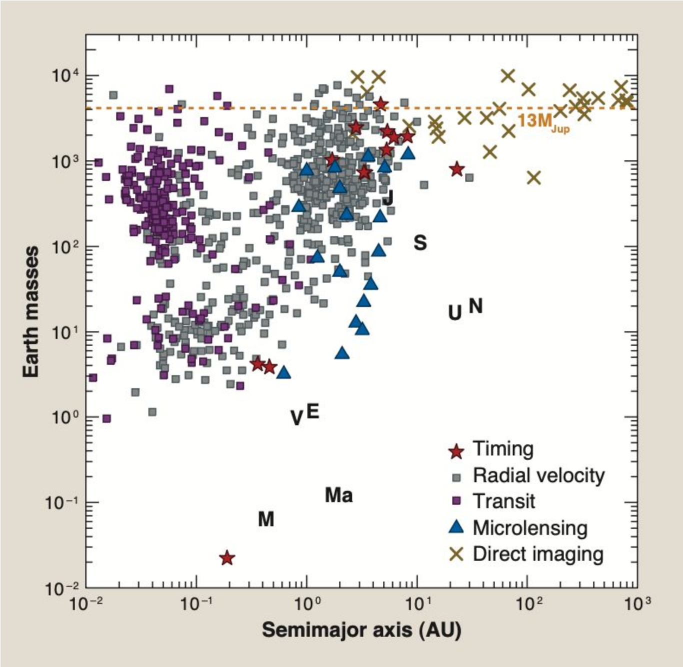


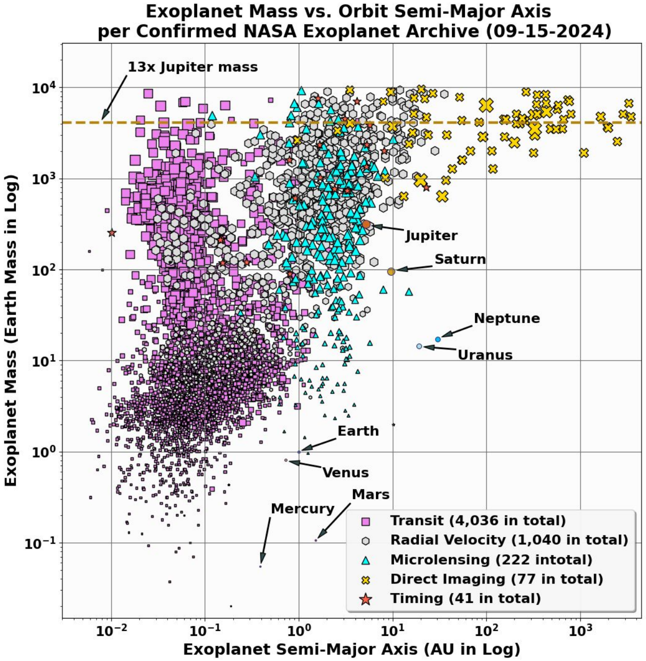


- Included solar system figures
- Newer methods (transit specifically) reveal a ton of new smaller planets, likely because of the nature of the method of discovery

New Paper

Prof. Seager's 2013 Paper

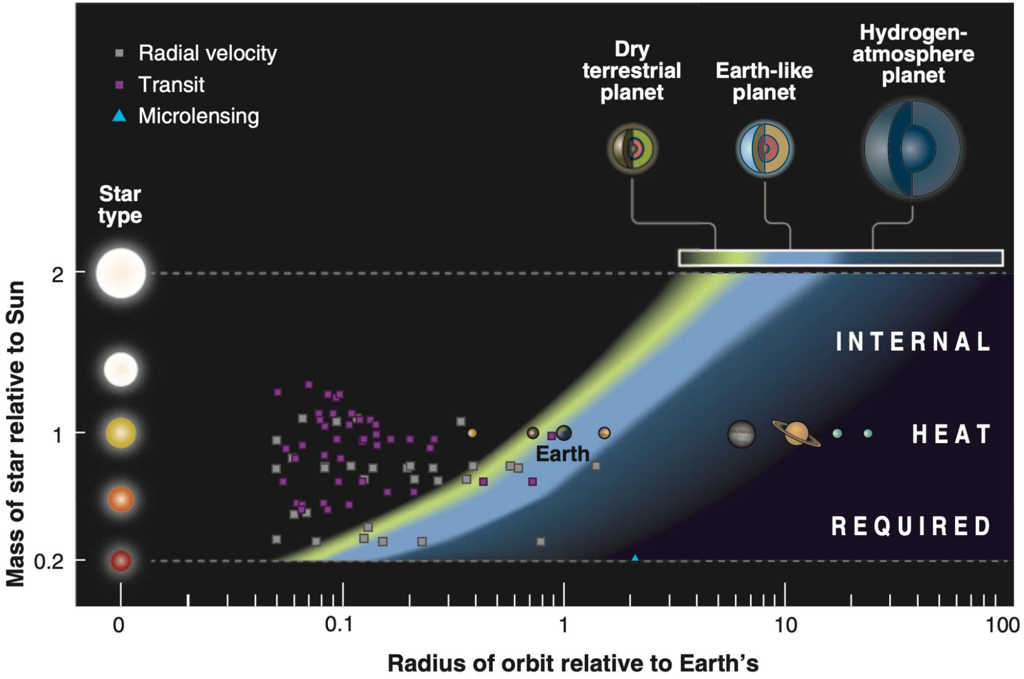




- Included solar system figures
- Newer methods (transit specifically) reveal a ton of new smaller planets, likely because of the nature of the method of discovery
- Solar system planets further than standard for their masses

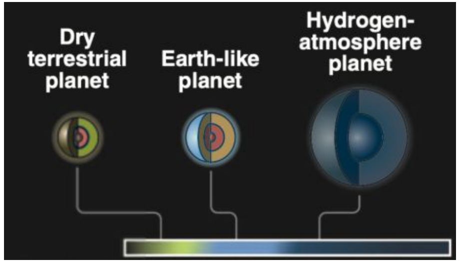
New Paper

Prof. Seager's 2013 Paper: Star Mass v. Exoplanet Orbital Relationship



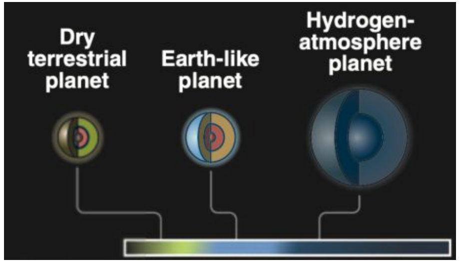
New Paper

Prof. Seager's 2013 Paper introduces 3 types of exoplanets:



New Paper

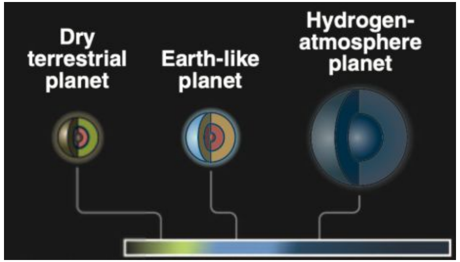
Prof. Seager's 2013 Paper introduces 3 types of exoplanets:



My new revision includes **4 types of exoplanets:**

New Paper

Prof. Seager's 2013 Paper introduces 3 types of exoplanets:



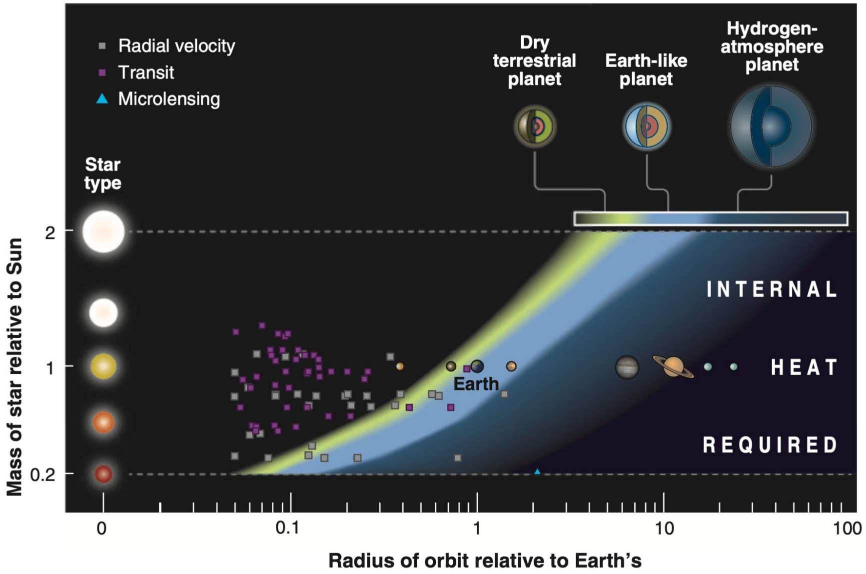
My new revision includes **4 types of exoplanets:**

- **Terrestrial:** planetary radius ≤ 1 Earth radius
- **Super-Earth:** 1 Earth radius < planetary radius ≤ 3.86 Earth radius
- **Neptune-Like:** 3.86 Earth radius < planetary radius ≤ 9.14 Earth radius
- **Gas-Giant:** 9.14 Earth radius < planetary radius

(based on NASA classifications)

New Paper

Prof. Seager’s 2013 Paper: Star Mass v. Exoplanet Orbital Relationship



https://doi.org/10.3847/2515-5172/add46f

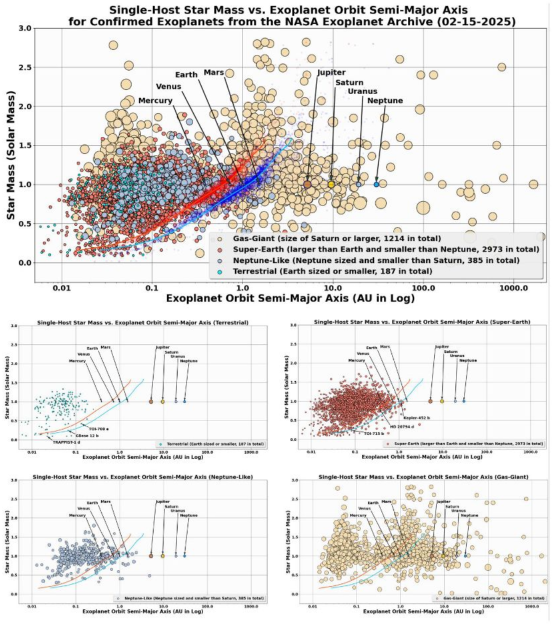


Figure 1. Top: single-host star mass vs. exoplanet orbital semi-major axis diagram. Exoplanets are colored by type and scaled by radius. Each pair of red and blue markers show the computed inner and outer HZ boundaries respectively for each host star, with the 5th-degree polynomial fits (from the least squares regression) indicated by the red and blue solid curves. Graphs on the mid-left, mid-right, bottom-left, and bottom-right are the breakdown diagrams for Terrestrial, Super-Earth, Neptune-Like, and Gas-Giant exoplanets respectively.

Future Work

Future Work

- Continuing research in exoplanet classification.
  - More work in stellar system classification and developing good categorizations for exoplanets.

- Working on applying my research to newer ML models.
  - Potentially looking at synthetic data? (ex. Gen 3 Bern Model)

Thank you!!