Statistical modelling of habitable zones of stars across the Harvard spectral classification (M-K-G-F-A-B-O)

Introduction

The habitable zone, also called the “goldilocks” zone is of great interest to astronomers and astrobiologists because it indicates the region in a solar systems where life is possible. Whether it be complex life or primitive life or no life at all, planets and moons located in the habitable zone most of the time have the conditions necessary for life. In this Internal Assessment, I will be investigating into the habitable zones of the stars across the Harvard spectral classification which is the most common classification used to categorize stars, to determine the common and different characteristics among the zones of each type, and whether they truly provide habitable conditions or not.

The habitable zones are considered as such because it is in these regions that the conditions for life are theorized to exist: such as the right temperature, orbital period, orbital resonance which determines whether a star is tidally locked or not: an important feature for the right weather to exist, amount of starlight, debris field in proximity, and so on.

I have developed a personal passion for astrophysics which causes me to delve into its topics and participate in its outreach to the public during my free time. For this internal assessment, I learned to implement statistical methods using a contemporary programming language to analyze stellar data and draw conclusions. Specific to the topic, I enjoyed doing extensive research into the topics of habitable zones, ran simulations in simulators to understand their significance, and studied research papers and articles to understand what makes that zone special.

Research Question

My research question would be **“To what extent do the characteristics of the habitable zones of stars vary across the stellar spectra?”** In detail, I will be calculating the habitable zones of approx. 120,000 different stars, classified under the Harvard spectral classification. I will be comparing these habitable zones to see how they differ in characteristics, mainly in the zone’s distance from the parent star. To do this, there are a certain set of calculations that can be done to calculate the habitable zone if one knows some particular parameters of the host star.

Background Knowledge

Personal motivation

“Why are we here? Where do we come from?”, were one of the very first words I heard from the mouth of my still-favorite physics professor, Brian Cox. For a long time, I’ve always gazed up at the stars and wondered the same question Just like every other stargazer looking into the vast cosmos does, “Are we alone?” The universe is mysterious, the universe is captivating, the universe is terrifying, but most of all, no matter if we choose to ignore it or not - the story of the universe is our story. Quoting the famous astrophysicist, Neil Degrasse Tyson, “We are stardust brought to life, then empowered by the universe to figure itself out—and we have only just begun.”

The fact that we, as living beings, haven’t found life anywhere else in the universe that we have discovered so far is itself, astounding and terrifying at the same time. Does this mean that we’re extremely rare? Or that we’re the first life in the universe? Or the last? Is there in fact life in the distant regions of the universe, but we’re just too far away to see it? Is it that any life in the universe eventually develops so much, that it wipes itself out or gets wiped out after a point, and we’re next in line for that? Or is it that the conditions for life to form are extremely rare - as rare as finding the exact copy of an atom in the same universe.

All these questions make us search for what is it that actually makes life possible, because the fact that we appear to be alone at the moment makes life very special. However, searching for life in the universe is not as straightforward as it looks because life, as far as we know, is fragile and delicate. Therefore, we look at a variety of seemingly unrelated factors to determine whether life can/would have developed somewhere, one of which is the habitability region of a star. Thus, this internal assessment has given me the opportunity to take part in this search for our cosmological cousins.

Theory

Definitions:

- **Distance:** Distance in astronomy refers to the distance to an object from the Earth’s instant position around the Sun’s orbit. Distance is mainly represented in three units, the **Parsec**, the **light year**, and the **Astronomical Unit**. The most common method of calculating distance to astronomical bodies is the parallax method/triangulation. The parallax method involves finding the shift in the observed body's position measured twice, i.e, every six months against the backdrop of interstellar space. The angle of shift is called the parallax angle, which is most commonly in arcseconds, and the formula for obtaining the distance (*in parsecs*) based on the parallax angle (*in arcseconds*) is given below:

$$d = \frac{1}{\tan{(p)}} \approx \frac{1}{p}$$

The approximation made in the above formula where $\tan(p) = p$, is valid for extremely small values such as arcseconds which are smaller than a degree by a factor of 3600th.

- **Luminosity**: Luminosity is considered to represent the quantity of power in astrophysical systems. It is the total amount of energy emitted by an astronomical body, such as a star or a galaxy in unit time. Its unit is the same unit as that of power - Joules per second or watts. Astronomical luminosity values can reach large values, so are usually represented in terms of luminosity of the sun (L_{sun}). The formula for luminosity L is given by the Stefan-Boltzmann law in terms of a star's surface area (which is $4\pi R^2$ because we assume stars are spherical), temperature T, and the Stefan-Boltzmann constant σ:

$$L = 4\pi R^2 \sigma T^4$$

- **Apparent Magnitude:** Apparent magnitude is a logarithmic measure of the apparent brightness of an astronomical body. Unlike the objective quantities such as luminosity and absolute magnitude, it is a subjective quantity that depends greatly upon the distance to the astronomical body in question, in that a closer but objectively dimmer body might have almost the same apparent magnitude value as a farther but objectively brighter body. It has an inverse relation, i.e, the lesser the value, the brighter the object (this is by virtue of it being a logarithmic scale). Apparent magnitude is calculated through flux photometry.

- **Absolute Magnitude:** Absolute magnitude is a logarithmic measure of the objective luminosity of an astronomical body, unlike the subjective apparent magnitude. Absolute magnitude doesn't depend on distance (it is in fact equal to the apparent magnitude of a body when viewed at a distance of 10 parsecs). It is an inverse scale just like apparent magnitude, with lesser values indicating higher luminosity. Absolute magnitude, however, differs based on which filter band is being considered for the astronomical body. The most common form, is the absolute visual magnitude, which uses the visual band of the spectrum (in the UBV photometric system). This absolute visual magnitude is represented by M_V, where the subscript represents the band filter being considered.

- **Absolute Bolometric Magnitude:** There is, in fact a more general form of the absolute magnitude called the Absolute Bolometric Magnitude, which takes into account electromagnetic radiation across all wavelengths, along with those that have been unobserved due to instrumental passband (the band of radiation that passes through without being detected), the Earth's atmospheric absorption, and interference by interstellar dust. Calculating the absolute bolometric magnitude involves applying a correction constant to the absolute visual magnitude. The correction constant is called the **Bolometric Correction Constant**, which is different for each spectral type in the Harvard spectral classification. Furthermore, the values of each correction constant has been long debated, and there exist many tables out there each with its values of constants.

- **Harvard spectral classification:** This is a classification[1] which categorizes stars based on their peak spectral wavelength and temperature into 7 different types (M-K-G-F-A-B-O). The temperature and wavelength ranges for each type are:
    - M-type (Temperature: < 3500 K, Wavelength: > 828 nm)
    - K-type (Temperature: 3500 K to 5000 K, Wavelength: 828 nm to 579 nm)
    - G-type (Temperature: 5000 K to 6000 K, Wavelength: 579 nm to 483 nm)
    - F-type (Temperature: 6000 K to 7500 K, Wavelength: 483 nm to 386 nm)
    - A-type (Temperature: 7500 K to 11,000 K, Wavelength: 386 nm to 263.5 nm)
    - B-type (Temperature: 11,000 K to 25,000 K, Wavelength: 263 nm to 116 nm)
    - O-type (Temperature: > 25,000K, Wavelength < 116 nm)

Methodology for calculation

Before calculating the inner and outer habitable radii, the absolute luminosity of the host star must be estimated:

1. Calculate the absolute visual magnitude of the host star based on the star's apparent magnitude:

[1] "Harvard Spectral Classification." *Physics and Universe*, 12 June 2013, physicsanduniverse.com/harvard-spectral-classification/. Date acc. 10th Dec., 2018

$$M_v = m_v - 5\log\left(\frac{d}{10}\right)$$

(M_v = absolute visual magnitude,

m_v = apparent visual magnitude,

d = distance to Earth in parsecs)

2. Calculate the bolometric magnitude of the host star:

$$M_{bol} = M_v + BC$$

(M_{bol} = Bolometric magnitude,

M_v = absolute visual magnitude,

BC = Bolometric correction constant [*The values for the constant differ based on the spectral type*[2]])

3. Calculate the absolute luminosity of the host star based on the above values:

$$\frac{L_{star}}{L_{sun}} = 10^{\left[\frac{M_{bolstar} - M_{bolsun}}{-2.5}\right]}$$

($\frac{L_{star}}{L_{sun}}$= Luminosity of the star with respect to the Sun's luminosity,

$M_{bolstar}$, = Bolometric magnitude of the star,

M_{bolsun} = Bolometric magnitude of the sun,

-2.5 is a constant value used for comparing stellar luminosities, which is known as "Pogson's Ratio")

After obtaining the luminosity of the star in terms of the sun, the approximate boundaries of the star's habitable zones can be calculated based on the following formulae:

$$r_i = \sqrt{\frac{L_{star}}{1.1}}$$

---

[2] Astronomical terms and constants, astro.princeton.edu/~gk/A403/constants.pdf. Date acc. 8th Sept., 2018.

$$r_o = \sqrt{\frac{L_{star}}{0.53}}$$

Where:

r_i = the inner boundary of the habitable zone in astronomical units (A.U).

r_o = the outer boundary of the habitable zone in astronomical units (A.U).

L_{star} is the absolute luminosity of the star.

1.1 and 0.53 = constant values representing stellar flux at the inner and outer radii respectively (based on Kasting et al., 1993[3]; Whitmire et al., 1996[2])

The approximate habitable zone radii has been calculated using stellar luminosity and stellar flux following using methods presented by the study of Whitmire et al.[4].

A sample calculation

Take the star HIP 2 (parallax: $20.85\,mas$, app. mag: 9.4017, B-V index: 0.999). We calculate the following:

The star's classification = K-type (extracted from table for B-V index 0.999)

The star's BC constant = -0.2 (extracted from correction table for K-type[5])

The star's distance to earth = $\frac{1000}{20.85} = 47.96\,pc$

The star's absolute magnitude = $9.4017 - 5\log(\frac{47.96}{10}) = 5.9978$

The star's bolometric magnitude = $5.9978 + (-0.2) = 5.7973$

The star's relative luminosity = $10^{[}\frac{5.7973 - 4.72}{-2.5}] = 0.3705 L_{\odot}$

The star's habitable zone's inner radius = $\sqrt{\frac{0.3705}{1.1}} = 0.5803\,AU$

---

[3] Kasting, James; Whitmire, Daniel; and Reynolds, Ray (1993). Habitable zones around main sequence stars. Icarus 101: 108-128. Date acc. 7th Sept., 2018.

[4] Whitmire, Daniel; Reynolds, Ray, (1996). Circumstellar habitable zones: astronomical considerations. In: Doyle, Laurence (ed.). Circumstellar Habitable Zones, 117-142. Travis House Publications, Menlo Park. Date acc. 17th Jan., 2018.

[5] Astronomical terms and constants, astro.princeton.edu/~gk/A403/constants.pdf. Date acc. 8th Sept., 2018.

The star's habitable zone's outer radius = $\sqrt{\frac{0.3705}{0.53}} = 0.8390\,AU$

And in this way, I use the programming language Python to calculate the inner and outer radii for 113942 filtered stars from the data set, statistically analyze, and graph the radii values in the Analysis section.

Dataset

The dataset I'll be using for this is from the famous Hipparcos catalogue[6], which is an official compilation of stellar data that is quite popular among the astronomical community. It has data from nearly 120,000 stars. There are multiple so-called Hipparcos datasets out there, but none of them are complete with the entire catalogue. However, I was able to procure an updated/revised version of the dataset from the VizieR archives, which is where the original catalogue was published.

The Hipparcos catalogue provides me the values for apparent visual magnitude, parallax, and B-V color indices which will be used for calculating the habitable zones. I will need to map B-V color index to the corresponding spectral type and obtain the bolometric correction constant for that specific type, for which the mapping table is in a paper published by Princeton university[5]. Furthermore, Hipparcos also provides me error correction values (absolute uncertainty), which I shall be using to calculate the erroneousness and uncertainty of my results.

Methodology for analysis

To extract, categorize, calculate, and visualize the data for the approx. 120,000 stars present, I will be using the Python programming language and its libraries -- which is commonly used in astrostatistics and data science in general.

To visualize the habitable zones, I have decided that a circular graph would be suitable, since it makes it easier to observe and study the radii's distributions. Matplotlib, which is the standard plotting library for Python, provides an option to plot circular and scatter graphs for the radii. For each spectral type, I plan to plot a separate graph and compare the resulting radii distribution of the inner radii and outer radii.

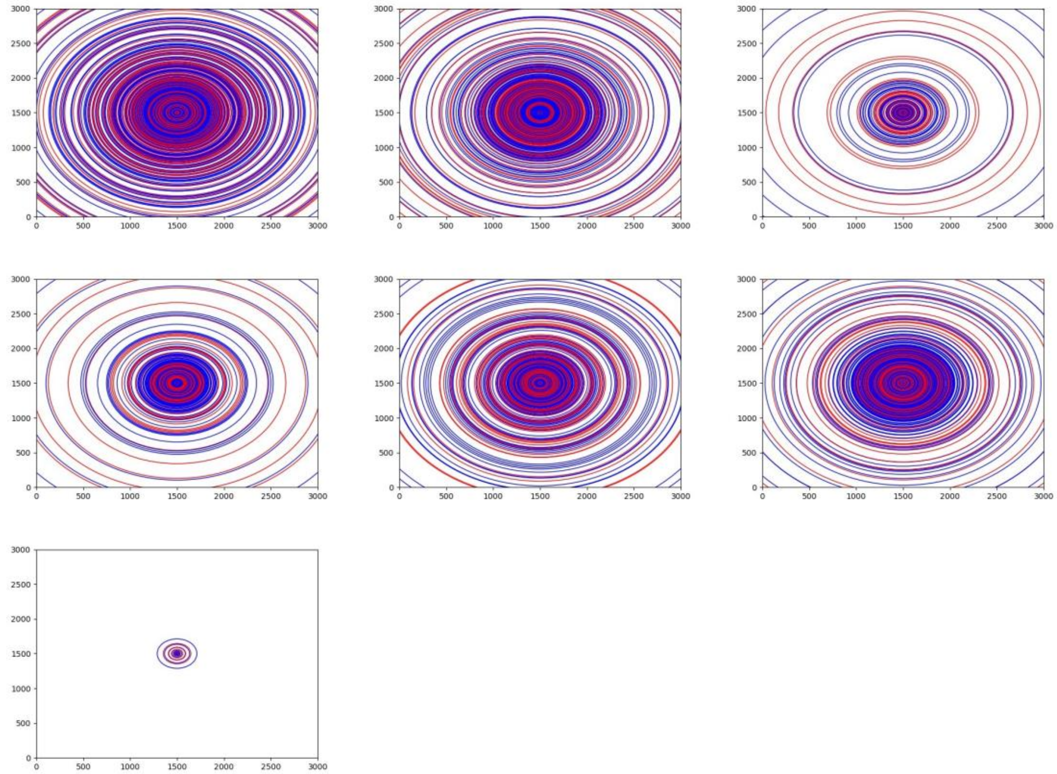
The Python code used for analysis and graphing:

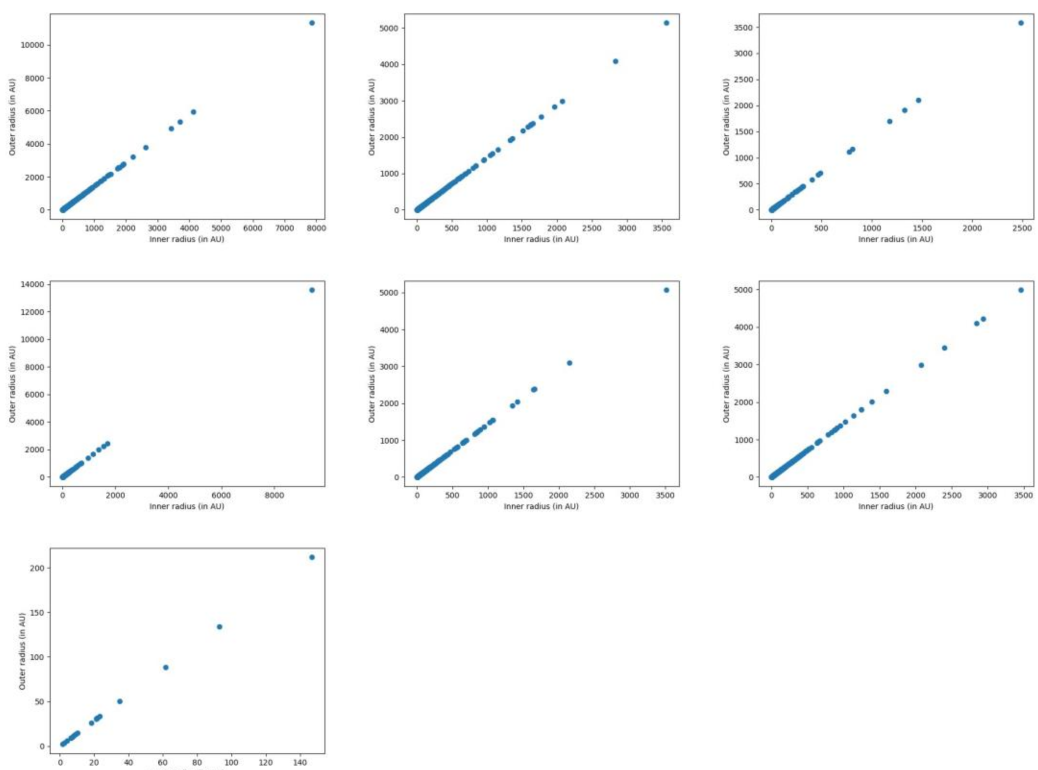
[6] VizieR {{CatName}}, cdsarc.u-strasbg.fr/viz-bin/Cat?cat=I%2F311. Date acc. 13th Aug., 2018.

1. newanalysis.py - [link]
2. calculations.py - [link]
3. uncertainties.py - [link]

Descriptive statistical analysis of data collected

Circle and line plots for the habitable zones of each star type (x-axis: in AU, y-axis: in AU):
(From left to right: M, K, G, F, A, B, O





Statistical analysis:

Stellar type	Inner radius mean (in AU)	Inner radius avg. uncertainty (in AU)	Outer radius mean (in AU)	Outer radius avg. uncertainty (in AU)	Inner radius median (in AU)	Outer radius median (in AU)	Inner radius std. deviation (in AU)	Outer radius std. deviation (in AU)
M	37.3200	5.4027	53.7651	16.1543	19.2451	27.7254	140.525	202.4473
K	10.0177	1.1331	14.4320	3.3881	5.9821	8.6182	46.0888	66.3979
G	3.3730	0.1331	4.8593	0.3980	1.2603	1.8157	29.7758	42.8966
F	4.3455	0.2805	6.2604	0.8387	2.2404	3.2276	58.5523	84.3533
A	11.1039	0.9566	15.9968	2.8603	5.5546	8.0023	51.0363	73.5254
B	35.7840	3.3817	51.5522	10.1114	19.4934	28.0832	99.3264	143.0946
O	26.3727	4.6162	37.9938	13.8027	14.3237	20.6354	35.0173	50.4477

(The reason for calculating median as well as mean is to analyze the symmetry of the distribution. If the mean is close to the median, that means the data points converge around a specific value indicating that there is a pattern to the data and it is not entirely random. Furthermore, mean values are affected by the presence of outliers in data points, as seen in the scatter plots. Outliers cause the mean to take on nonsensical values. Therefore, we also consider the median values because the nature of median is such that it is not affected by outliers.)

Analysis of results

In total, there were 9184 M-type stars, 33623 K-type stars, 16642 G-type stars, 30414 F-type stars, 17484 A-type stars, 6575 B-type stars, and 20 O-type stars. 113942 stars were analyzed and the remaining 6058 out of the 120,000 were filtered out due to erroneous values.

For M-type and K-type stars, graphical and statistical analysis gives a mean inner and outer radius of 37.32 and 53.76 AU, and 10.01 and 14.43 AU respectively. This is quite unusual, as these stars are not bright enough for their habitable zones to be that far out. I was expecting the radii to be under an AU, but clearly not. An explanation for this is the fact that M and K types are known to be flare stars. They undergo periodic radiation outbursts that cause a spike in their luminosity for a brief time, thus accounting for the large radii values. This hunch is further reinforced by the fact that the mean inner and outer standard deviations for M-type are 140.52

and 202.44 AU and the inner and outer uncertainties are the largest out of all, indicating an expected large variation in the data.

For G-type stars, statistical analysis gives a mean inner and outer radius of 3.37 and 4.85 AU. These values are expected, since research has shown that the Sun (which is a G-type) has the habitable zone lying between 0.97 to 1.37 AU[7]. Furthermore, the median inner and outer values are 1.26 and 1.81 AU which is close to the range of our Sun. The inner and outer radii uncertainty values are also feasible enough to fit the description of sun-like stars.

For the next four types: F, A, and B, the mean inner and outer radii increases in range and value, which is expected since each type is hotter than the previous one, and rising temperature corresponds to rising luminosity and thereby increasing radii values. The inner and outer radii uncertainty values also increase steadily as the hot stars of these types vary greatly in luminosity due to certain physical mechanisms. However, for O, which is after B, the inner and outer radii values decrease. This is due to the fact that only 20 O-type stars were filtered during analysis, which I suspect is because there is no BC correction value for O4 and above[8].

Now, to compute the size of the habitable zones, we subtract the inner and outer radii. For the 7 types, the zone sizes (in AU) are: M-type: mean = 16.44, median = 8.48, K-type: mean = 4.41, median = 2.63, G-type: mean = 1.48, median = 0.55, F-type: mean = 1.91, median = 0.98, A-type: mean = 4.89, median = 2.44, B-type: mean = 15.76, median = 8.58, O-type: mean = 11.62, median = 6.31. Not considering M and K because of their flaring nature, We see that the zone is smallest for G-type and F-type stars and largest for A-type and B-type stars (O-type's data is not being considered due to the lack of a BC constant). This is physically accurate because rocky planets that have surfaces for liquids water are often formed in the region of G and F types' habitable zones, thus leading to possibly planets with life which we have speculated for quite some time now[9]. However, while A and B's zones are bigger in size, the planets that are formed in their habitable zones are gas giants and the sort, which are not habitable for life unless they have rocky moons with water.

Conclusion

---

[7] "Circumstellar Habitable Zones." *Circumstellar Habitable Zones - Habitable Zones - NAAP*, astro.unl.edu/naap/habitablezones/chz.html. Date acc. 21st Oct., 2018.

[8] Astronomical terms and constants, astro.princeton.edu/~gk/A403/constants.pdf. Date acc. 8th Sept., 2018.

[9] "What's the Recipe for a Habitable Exoplanet?" EarthSky, earthsky.org/space/nasa-nexss-recipe-for-habitable-exoplanets. Date acc. 8th Jan., 2019

Thus, by using statistical methods, I was able to compute the habitable zone radii distribution of the various star types of the Harvard spectral classification, and make conclusions based on the results of the analysis while also accounting for anomalies. This is a testament to the power of mathematical methods in analyzing physical properties. In summary, my curiosity in planetary habitable zones combined with my programming skills have enabled me to successfully carry out this investigation to come to a personal consensus about where life is possible in stars after all.

However, although we have computed the distribution of habitable zones, the values represent the effective habitable zones but not the actual zones where the conditions are present for life to be possible. To actually determine the habitable zone, other factors such as peak wavelength of starlight, asteroid proximity, gaseous material distribution in the solar system, effective temperature of a planet at that distance, and so on must be taken into account, after which the actual zones can be determined which could turn out to be quite small when compared to the empirically determined effective zones.

Some limitations (and thereupon possible extensions) of this investigation are:

- The method for calculating habitable zones has been derived statistically - the values 0.53 and 1.1 which were used to calculate the inner and outer radii from the luminosity value aren't arbitrary but have been derived statistically from analyzing a large amount of stars and the properties of their known planets.
    - A possible extension to the investigation to overcome this limitation is to compute the habitable zone radii using a different method - ideally using a formula or the like to arrive at different and possibly more accurate results.
- The method for calculating habitable zones depends on the luminosity of the star, which is not a stable value due to which it isn't an accurate measure of the habitable zones of all stars, some exceptions being flare (M-type) stars.
    - A possible extension to the investigation to overcome this limitation is to use a different method that either does not involve luminosity (not likely) or uses luminosity and a couple more quantities to compute the habitable zone.
- Not many O-type stars were filtered and analyzed since the bolometric correction constant table that I have used lacks the BC correction value for O4 and above, due to which I had to make the minor adjustment in the analysis code to use the O5's BC correction value for all of the O subtypes.
    - A possible extension to the investigation to overcome this limitation is to use a different correction constant table which contains the BC correction values for the O subtypes.

Bibliography

1. Research papers:
    a. Fogg, Martyn J., (1992). An estimate of the prevalence of biocompatible and habitable planets. *Journal of the British Interplanetary Society*, 45: 3-12. Date acc. 9th Sept., 2018.
    b. Habets, G.M.H.J.; Heintze, J.R.W., (1981). Empirical bolometric corrections for the main sequence. *Astron. Astrophys. Suppl. Ser.* 46: 193-237. Date acc. 11th Sept., 2018.
    c. Kasting, James; Whitmire, Daniel; and Reynolds, Ray (1993). Habitable zones around main sequence stars. *Icarus* 101: 108-128. Date acc. 7th Sept., 2018.
    d. Kasting, James F. (1996). Habitable zones around stars: An update. In: Doyle, Laurence (ed.). *Circumstellar Habitable Zones*, 117-142. Travis House Publications, Menlo Park. Date acc. 21th Oct., 2018.
    e. Lang, Kenneth (1992). *Astrophysical Data: Planets and Stars*. Springer-Verlag, New York. Date acc. 10th Nov., 2018.
    f. Whitmire, Daniel; Reynolds, Ray, (1996). Circumstellar habitable zones: astronomical considerations. In: Doyle, Laurence (ed.). *Circumstellar Habitable Zones*, 117-142. Travis House Publications, Menlo Park. Date acc. 17th Jan., 2018.
2. Articles, documents and web links:
    a. Dataset: VizieR {{CatName}}, cdsarc.u-strasbg.fr/viz-bin/Cat?cat=I%2F311. Date acc. 13th Aug., 2018.
    b. Constants and conversions: Astronomical terms and constants, astro.princeton.edu/~gk/A403/constants.pdf. Date acc. 8th Sept., 2018.
    c. “Circumstellar Habitable Zones.” Circumstellar Habitable Zones - Habitable Zones - NAAP, astro.unl.edu/naap/habitablezones/chz.html. Date acc. 21st Oct., 2018.
    d. “Harvard Spectral Classification.” *Physics and Universe*, 12 June 2013, physicsanduniverse.com/harvard-spectral-classification/. Date acc. 10th Dec., 2018
    e. “What's the Recipe for a Habitable Exoplanet?” EarthSky, earthsky.org/space/nasa-nexss-recipe-for-habitable-exoplanets. Date acc. 8th Jan., 2019