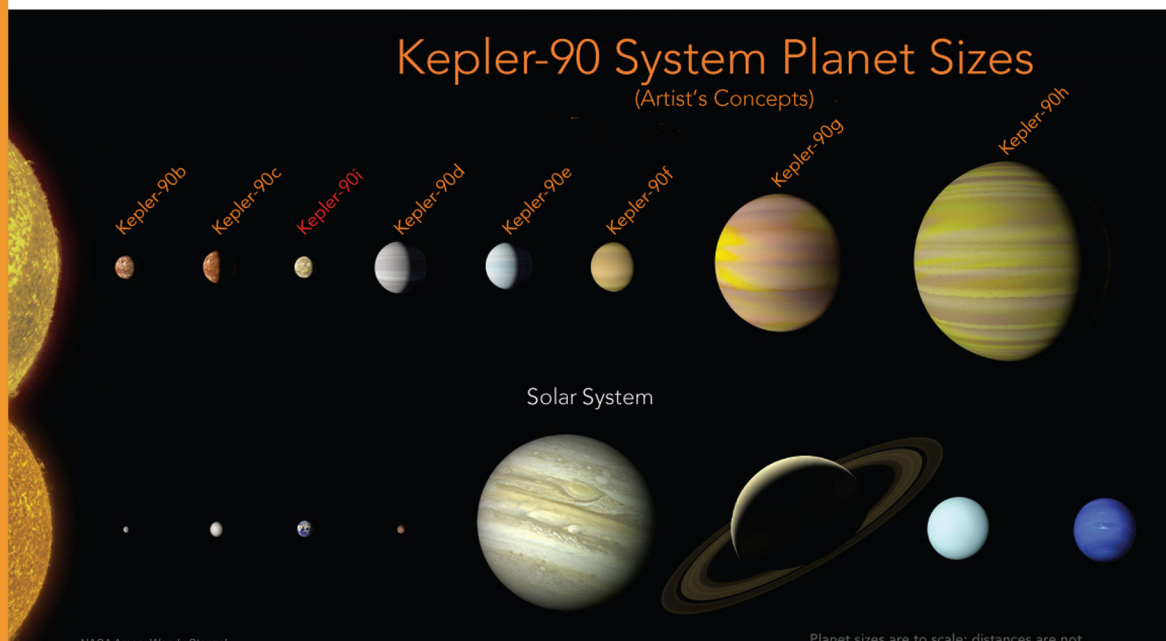
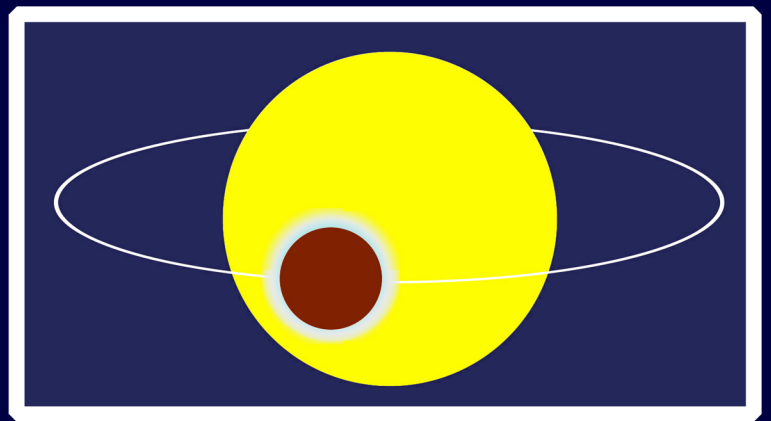


Exoplanetas: minería con datos reales, de aplicación en el aula

Joaquín Álvaro Contreras



EXOPLANETAS: MINERÍA CON DATOS REALES, DE APLICACION EN EL AULA

Joaquín Álvaro Contreras

Publicaciones de ApEA

**Asociación para la Enseñanza de la Astronomía
Nº 43 (Secundaria)- Abril 2023**

Publicaciones de ApEA nº 43
Asociación para la Enseñanza de la Astronomía

© Texto: Joaquín Álvaro Contreras

© Dibujos e imágenes: Joaquín Álvaro Contreras, NASA.

Dibujo de Portada: Tanguy

Foto de Portada: Sistema planetario de Kepler 90 Ilustración de Wendy Stenze, NASA Ames.

Comité de Redacción:

Ederlinda Viñuales, Antonio Arribas y Ricardo Moreno

Dirección:

Ricardo Moreno (Vocal Editor de Publicaciones de ApEA)

rmluquero@gmail.com

ISBN: 978-84-942304-4-8



Esta obra está licenciada bajo la Licencia Creative Commons Atribución 4.0 Internacional. Para ver una copia de esta licencia, visita <http://creativecommons.org/licenses/by/4.0/>.

EL AUTOR

Joaquín Álvaro Contreras



Es Licenciado en Ciencias Físicas, especialidad de Astrofísica, por la Universidad Complutense de Madrid (UCM). Hizo el programa de Doctorado en Física Teórica y Física Automática y Robótica. Ha sido investigador sobre radiación cósmica en el Instituto Nacional de Técnica Aeroespacial, Grupos de la Comisión Nacional de Investigación del Espacio (INTA-CONIE) en la década de los 80, con Juan Sequeiros y José Medina. Durante 14 años ha sido docente en la UNED, (1985-1999).

Su actividad profesional se ha movido en el campo de Informática, (grandes sistemas). Además, ha mantenido una actividad dilatada como divulgador de la Astronomía y de la Física: cursos para docentes, conferencias, etc. Es socio fundador y presidente actual de AstroCuenca. También tiene experiencia en diseño museístico del área de Astronomía en el Museo de las Ciencias de Castilla La Mancha. Es miembro de la *Sociedad Española de Astronomía* (SEA), de la *European Astronomical Society* (EAS) y de la *Europlanet Society* (EPS).

Actualmente es Presidente de la Federación de Asociaciones Astronómicas de España (FAAE), miembro de la Comisión Pro-Am de la SEA y Miembro del *Regional Hub for Spain and Portugal of the Europlanet Society*.

ÍNDICE

Autor.....	3
Presentación.....	7
1ª Parte	
Introducción.....	9
El fichero de datos.....	11
Importación a una hoja de cálculo.....	12
Primer ejercicio: Método de detección.....	12
Segundo ejercicio: Distribución espacial	15
Tercer ejercicio: Distancia a la Tierra	17
Anexo 1.....	20
2ª parte	
Sistemas planetarios más allá del Sistema solar:	25
a) distancia a la estrella	26
b) Zona de habitabilidad	28
c) Tipo de estrella.....	30
Anexo 2.....	32
3ª parte	
Relación Masa-Radius.....	35
Distribución de densidades.....	38
El algoritmo de ‘Chen & Kipping.....	39
Modelos mejorados. Alternativas.....	42
Exoplanetas e Inteligencia Artificial.....	43
Conclusión.....	44
Anexo 3.....	45

PRESENTACIÓN

Hoy día, la inmensa cantidad de información recopilada por los grandes observatorios y los telescopios espaciales plantea un reto importante para los investigadores: el ingente volumen de datos que se va acumulando supera ampliamente la capacidad de análisis por parte de la comunidad científica. Por otra parte, al menos en la Astronomía, se ha entendido que esta información es un recurso que va más allá de la propiedad exclusiva de los responsables de los proyectos que la han generado, incluidas las Agencias Espaciales, que han hecho posible el desarrollo de la costosa instrumentación con la que se explora el universo.

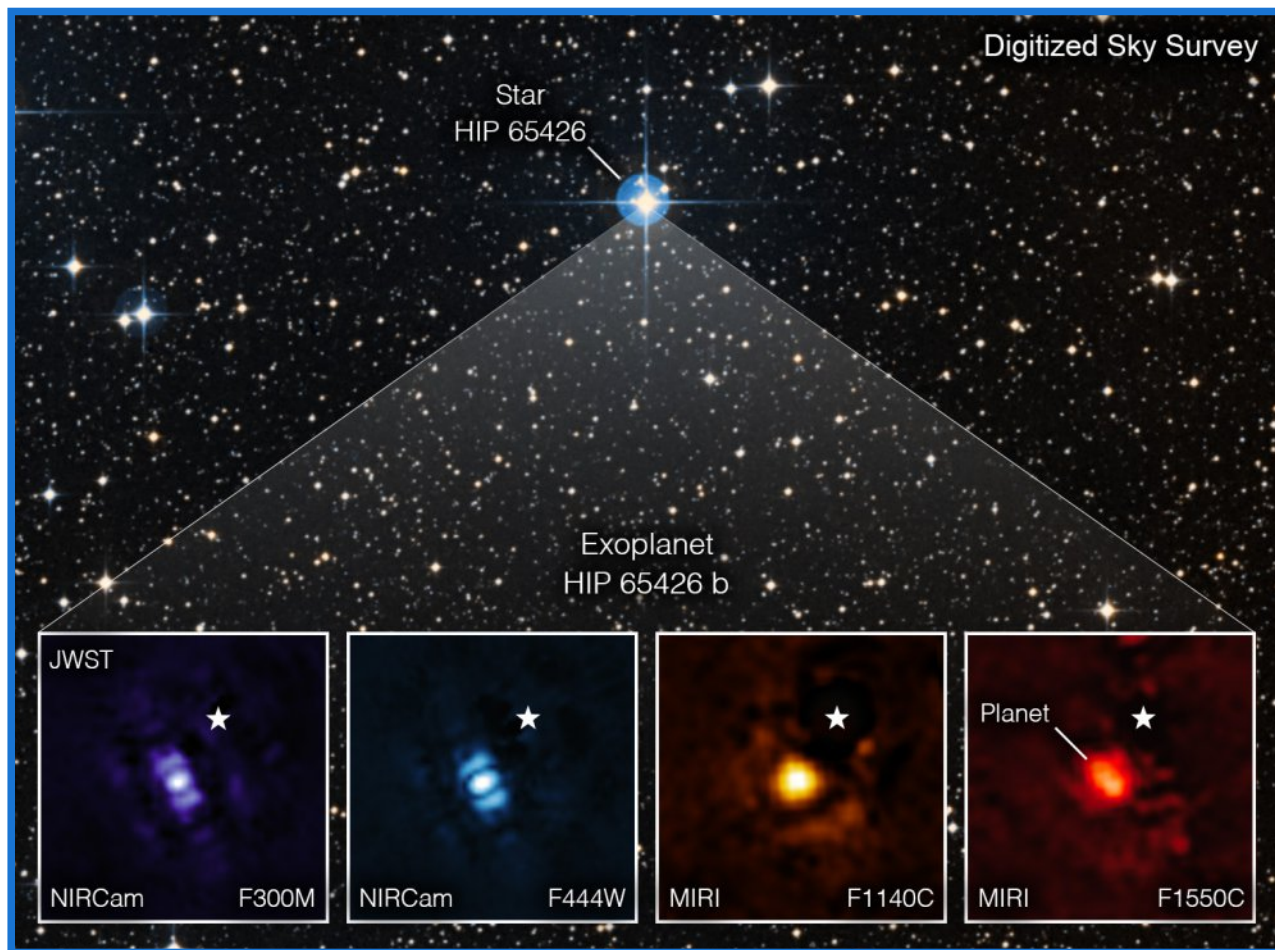
De esta manera, hace ya algún tiempo que los datos, de gran calidad, que constantemente se van añadiendo a esta base de conocimiento observacional, se comparten y están disponibles en ‘abierto’ tanto para los científicos como para cualquiera que quiera colaborar en su estudio y análisis.

Es cierto que, en la mayoría de los casos, no es fácil interpretar esta información si no se dispone de una formación adecuada. En definitiva se trata de ‘hacer ciencia’ a partir de datos empíricos y esto requiere un entrenamiento adecuado. Aunque los datos se ofrecen en general ya calibrados, y se estructuran en catálogos con herramientas de acceso *ad hoc*, siempre es necesario un proceso de aprendizaje que va desde el conocimiento del significado de cada parámetro, hasta la maduración de los objetivos que se persiguen tratándolos, a lo que deben seguir las correspondientes tareas de análisis, algo a lo que llamamos ‘minería de datos’.

En esta publicación se pretende hacer una incursión sencilla en este ámbito, presentando un catálogo asequible de los exoplanetas conocidos, a partir del cual se pueden realizar diversos ejercicios en un nivel posible para los estudiantes de secundaria, y que no requieren herramientas complicadas. El objetivo es que este intento pueda servir de toma de contacto con ‘*el observatorio virtual*’, motivar al alumnado y adquirir hábitos en el método científico.

El trabajo desarrollado en estos tres artículos forma parte de un proyecto llamado ‘Museos Ampáticos’ que, bajo la tutela del Museo de las Ciencias de Castilla La Mancha, estaba previsto llevar a cabo con alumnos de altas capacidades de los centros educativos de la Región a lo largo de un curso escolar. La irrupción de la pandemia, justo cuando iba a dar comienzo este programa, postpuso *sine die* su desarrollo, junto con otros proyectos similares en otras áreas de competencias, por lo que no podemos ofrecer resultados del deseable *feedback* que debería haber acompañado al proyecto.

1ª PARTE



Esta publicación sobre *exoplanetas* tiene como propósito adentrarse de manera sencilla en el acceso a un conjunto de datos, para ver qué y cómo hacer con ellos, y mostrar desde el principio posibles trabajos a desarrollar en el aula con estudiantes de secundaria. La riqueza, abundancia y calidad de los datos originales admite desde un tratamiento sencillo, con simples hojas de cálculo Excel, hasta otras técnicas más sofisticadas con programación propia e inteligencia artificial de por medio, o bien directamente el uso de las herramientas ya disponibles en el VO (*Virtual Observatory*).

Introducción

Desde hace un tiempo, ya dilatado, los datos recopilados por los observatorios astronómicos profesionales, las misiones espaciales y hasta algunos propios de los amateurs, (aunque en menor grado), están

disponibles y accesibles de manera pública, y en general con grandes facilidades de consulta y descarga. En los últimos años el esfuerzo por integrar esta información en grandes bases de datos, compatibles y conectadas entre sí, ha sido importante y se

han creado paralelamente herramientas para facilitar tanto su acceso como su integración y explotación. A esto se le viene llamando Observatorio Virtual (VO).

El VO es una iniciativa de carácter internacional que nació en el año 2000 y cuyo objetivo es optimizar la explotación científica de los cientos de miles de datos existentes en los archivos astronómicos. La Alianza del Observatorio Virtual Internacional (IVOA) engloba una veintena de países, entre ellos España desde el 2004, (véase: [Spanish Virtual Observatory \[SVO\]](#)).

Iniciamos con este artículo una serie de publicaciones sucesivas sobre dos temas, elegidos a modo de ejemplo, el primero de ellos sobre *exoplanetas* y el segundo sobre *cúmulos estelares abiertos*, cuyo propósito es adentrarse de manera sencilla en el acceso a estos conjuntos de datos, qué y cómo trabajar con ellos, y mostrar desde el principio posibles trabajos a desarrollar en el aula con estudiantes de secundaria. La riqueza, abundancia y calidad de los datos originales admite desde un tratamiento sencillo con simples hojas de cálculo Excel, hasta otras técnicas más sofisticadas con programación propia e inteligencia artificial de por medio, o bien directamente el uso de las herramientas ya disponibles en el VO.

Aquí intentaremos ceñirnos a la opción más simple, con Excel, y sólo al final se propondrán soluciones más avanzadas a problemas de mayor alcance. Desde el principio se verá el carácter formativo y motivador del trabajo con datos reales como los utilizados en la investigación astronómica.

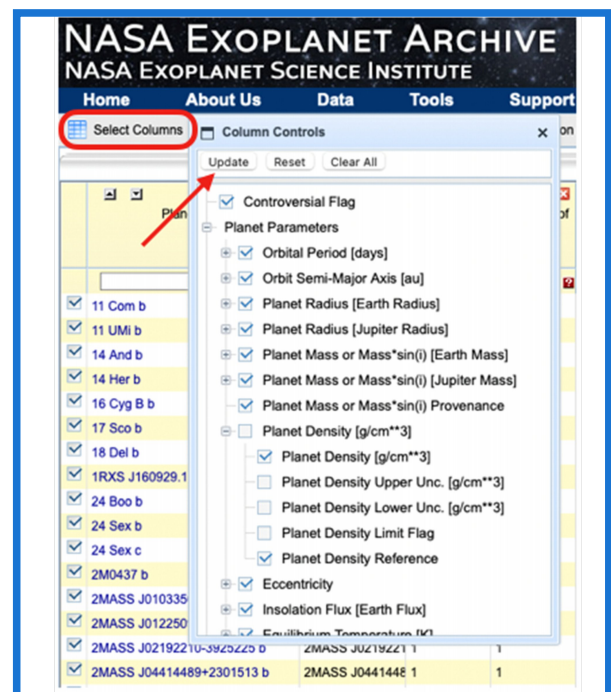
El fichero de datos

Para este primer trabajo utilizaremos el fichero¹ “[Planetary Systems Composite Data](#)” que contiene, a fecha del 9 de mayo de 2022, información muy detallada de los 5.030 exoplanetas confirmados hasta esta fecha. En la [página de NASA](#) por la que se accede a este fichero también hay otros conjuntos de datos así como la documentación pertinente relativa a los mismos.

a) Empezaremos por acceder al fichero² y descargarlo completo en formato .CSV

<https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ExoTbls&config=PSCompPars>

b) Una vez en la página ya se tiene una visión de este conjunto de datos, pero antes de proceder a su descarga vamos a seleccionar algunas columnas adicionales que no están incluidas en la versión ‘por defecto’ que se presenta aquí y que vamos a necesitar posteriormente.



[1] "This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program".

[2] En [NASA Exoplanet Archive](#) hay dos archivos: *Planetary Systems* y *Planetary Systems Composite Data*. La diferencia entre ellos es que el segundo integra datos en un único registro para cada exoplaneta recogidos de diferentes fuentes, mientras que ‘Planetary Systems’ puede incorporar varios registros para un mismo exoplaneta a partir de referencias diferentes. Haremos uso aquí del fichero compuesto: ‘Planetary Systems Composite Data’.

Para ello se *pincha* en '*Select Columns*' y en la ventana emergente, además de los parámetros ya marcados, que mantendremos, seleccionaremos los siguientes:

Planet Radius Reference

(dentro de Planet Radius [Earth])

Planet Density

Planet Density Reference

(dentro de Planet Density)

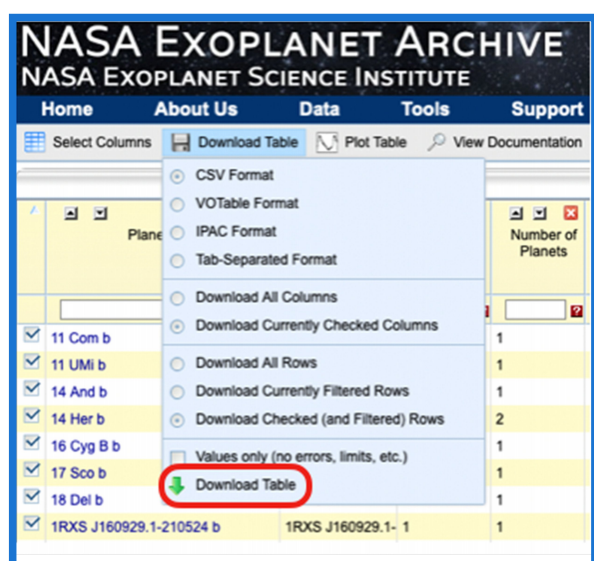
y de los Parámetros estelares:

Stellar Luminosity

Stellar Age

Podemos aprovechar aquí para eliminar de la selección otras columnas que no nos harán falta, como las magnitudes fotométricas estelares, (excepto la de Gaia, que sí utilizaremos), y los diferentes '*flags*'. Pero esto puede hacerse con tranquilidad más tarde, una vez importado el fichero a Excel .

Una vez seleccionados los datos que queremos descargar es preciso *pinchar* la opción '*Update*' y ya tendremos la página con la tabla según la hemos configurado lista para ser descargada. Y la descargamos.



c) Como se ha podido comprobar, no es necesario ningún registro/login previo para tener acceso y proceder a su descarga. Se recomienda hacer mención con el '*agradecimiento*' reseñado en el pie de la página anterior en cualquier trabajo que se publique derivado de estos datos.

d) Lo primero que debe hacerse con este fichero es abrirlo con un editor de texto y explorar mínimamente la información que contiene. Como podrá comprobarse, antes de los datos en sí hay una cabecera con la referencia de contenidos en cada columna. Es importante tener este repositorio siempre presente para consultarlo cuando sea necesario, especialmente en lo relativo a las unidades en las que se expresan los respectivos parámetros. (Se recomienda hacer una copia del fichero descargado antes de proceder a cualquier manipulación posterior).

e) La simple inspección de esta cabecera ya nos da idea de la cantidad de información que contiene, tanto de los exoplanetas en sí mismos como de las estrellas anfitrionas. A medida que vayamos haciendo uso de estos contenidos iremos explicando su significado.

Importación a una hoja de cálculo Excel.

a) En la exploración inicial, se habrá comprobado que en el fichero descargado el delimitador de columnas es una coma ',' y el separador decimal un punto '.'. Como lo habitual será tener configurado Excel con coma ',' como separador decimal, antes de importar el fichero .csv a una hoja de cálculo conviene preparar dicho fichero convenientemente para una correcta importación.

Se abrirá el fichero con un editor de texto sencillo (no Word o similar) y se procederá a:

- eliminar los registros de cabecera, (recuérdese haber hecho una copia previa del fichero original; también puede aprovecharse para guardar estos registros de cabecera con la información del contenido en un fichero independiente para tenerlo fácilmente disponible).

- reemplazar las comas ‘,’ por punto y coma ‘;’. Se utilizará este carácter ‘;’ como delimitador de campos (columnas).

- reemplazar los puntos ‘.’ por comas ‘,’. De esta manera el separador decimal será entendido perfectamente por nuestro Excel.

- guardar el fichero así modificado en el mismo formato .csv.

b) Ya es momento de importar el fichero .csv modificado a una tabla Excel. Puede importarse igualmente el fichero con los datos de cabecera a una pestaña específica del mismo fichero Excel con objeto de tenerla disponible con facilidad.

Para la importación de la tabla de datos sólo es preciso tener en cuenta el carácter delimitador con que se ha dejado preparado el .csv, en nuestro caso el punto y coma ‘;’.

Después de comprobar que la importación ha sido correcta, que los campos en cada columna están debidamente alineados y que los valores numéricos se han interpretado correctamente, es momento de hacerla lo más legible posible dando a cada columna el ancho necesario, inmovilizando paneles, al menos la primera fila con los nombres de las columnas y deseable también la primera columna con el nombre del exoplaneta. Como ya debe ser patente en este momento, las columnas con parámetros específicos relativos a los planetas se identifican con el prefijo ‘pl_’ mientras que los relativos a la estrella anfitriona del sistema planetario empiezan por ‘st_’ o ‘sy_’.

También ahora se puede prescindir de algunas columnas que no nos van a ser necesarias, como las referencias ‘disc_facility’ y los diferentes ‘flags’ incorporados al fichero, columnas que acaban en ‘...lim’. Mantendremos las columnas con nombre acabado en ‘...err1’ y ‘...err2’, que son las respectivas estimaciones de error (+, -) de los parámetros a que hacen referencia.

El número de registros que componen la hoja de cálculo finalmente dependerá de la fecha en que se haya realizado la descarga, (el fichero original de NASA se actualiza una o dos veces al mes). Para la fecha del 9/5/2022, este número es de 5.030 exoplanetas.

Primer ejercicio: Método de detección de exoplanetas

Este ejercicio puede servir para desarrollar una amplia temática en clase sobre los [métodos de detección de exoplanetas](#), (básicamente todos son métodos indirectos, excepto el de ‘imaging’), los procesos que lo hacen posible en cada caso y nos servirá también para explicar los sesgos instrumentales derivados de los procedimientos de detección, que se harán notorios un poco más adelante.

a) Utilizaremos para este ejercicio sólo las columnas ‘discoverymethod’ y ‘disc_year’, esto es método de detección y año del mismo. Se aconseja copiar el contenido de estas columnas a una pestaña nueva para hacer más fácil las sucesivas manipulaciones a las que se someterán estos datos. Puede aprovecharse también para cambiar el nombre de las cabeceras de las columnas si se lo considera conveniente, aunque no es lo más aconsejable, para no perder la homogeneidad con el repositorio de referencias.

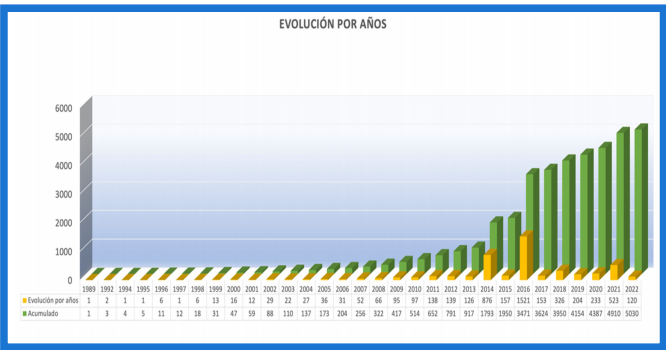
El objetivo de este ejercicio es ir familiarizándose con los datos contenidos en el fichero y tomar contacto con el tema ‘exoplanetas’, las dificultades instrumentales de detección, los principales métodos que lo hacen posible y un poco de la historia de esta nueva y pujante rama de la Astronomía.

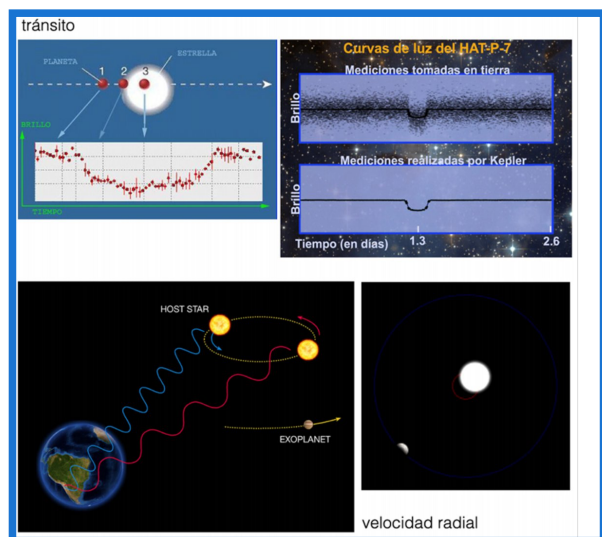
Se proponen tres simples análisis para los cuales habrá que hacer diferentes ordenaciones de estos datos: por año de descubrimiento y/o por método. En cada caso se pueden consolidar resultados y/o hacer agrupamientos según se considere necesario, para obtener tablas adicionales que puedan ser representadas gráficamente de manera fácil. Véase la tabla siguiente:

discoverymethod	núm.	%
Transit	3854	76,60%
Radial Velocity	927	18,40%
Microlensing	130	2,60%
Imaging	59	1,20%
Transit Timing Variations	22	0,40%
Eclipse Timing Variations	18	0,40%
Orbital Brightness Modulation	9	0,20%
Pulsar Timing	7	0,10%
Pulsation Timing Variations	2	0,00%
Astrometry	1	0,00%
Disk Kinematics	1	0,00%
	5030	100,00%

La detección de exoplanetas era algo inimaginable hasta hace relativamente poco tiempo. Siempre se había pensado que nuestros instrumentos ópticos estaban lejos de tener la capacidad suficiente para observar planetas fuera de nuestro Sistema Solar. La muy baja luminosidad de los mismos y su proximidad a la estrella que orbitan hacen que estos queden enmascarados por la luz de la estrella anfitriona. Hasta que se depuraron las técnicas de detección indirecta.

La siguiente gráfica muestra esta historia. Desde el primer exoplaneta descubierto en 1989 sólo han pasado algo más de 30 años, y sólo en los últimos 10 o 12 el ritmo de detección ha crecido de manera significativa.





En ambos casos la detección requiere una combinación de dos cosas:

a) Una sensibilidad manifiesta en los instrumentos de medida, esto es, fotometría hasta el orden de milésimas de magnitud aparente para detectar el eclipse de luz estelar durante el tránsito y variaciones mínimas en los corrimientos espectrales de dicha luz para el método de velocidad radial.

b) Una afortunada disposición del plano orbital del exoplaneta, de manera que dicho plano esté convenientemente orientado en la dirección de observación.

Estas dos técnicas tienen a su vez condicionantes adicionales. En el caso del método de velocidad radial, será más fácil la detección de planetas muy masivos que sitúen el centro de masas del sistema notoriamente fuera del centro de la estrella anfitriona, haciendo así posible la observación espectroscópica del movimiento de la estrella en torno al centro de masas.

En el caso del método del tránsito, también será más fácil detectar el oscurecimiento mínimo de la estrella

provocado por el eclipse de su luz si el planeta es más grande. Por otra parte, dado que se requieren al menos tres tránsitos sucesivos detectados para establecer el periodo de rotación del exoplaneta y confirmar así su descubrimiento, la técnica sólo estará dando resultados positivos, por ahora, con planetas de periodos orbitales cortos (próximos a la estrella).

A pesar de estas dificultades la sensibilidad actual de los equipos de aficionados y su saber hacer en mediciones fotométricas, han hecho posible que los amateurs estén detectando tránsitos de exoplanetas con equipos supuestamente modestos. También es posible colaborar en estas tareas de detección/confirmación analizando curvas de luz de los datos obtenidos por los telescopios espaciales. Un buen ejemplo es el [proyecto K2OJOS](#), desarrollado por miembros de la Asociación Astronómica Asturiana OMEGA en colaboración con el Instituto de Ciencias y Tecnologías Espaciales (ICTEA) de la Universidad de Asturias, a partir de datos de Kepler/K2.

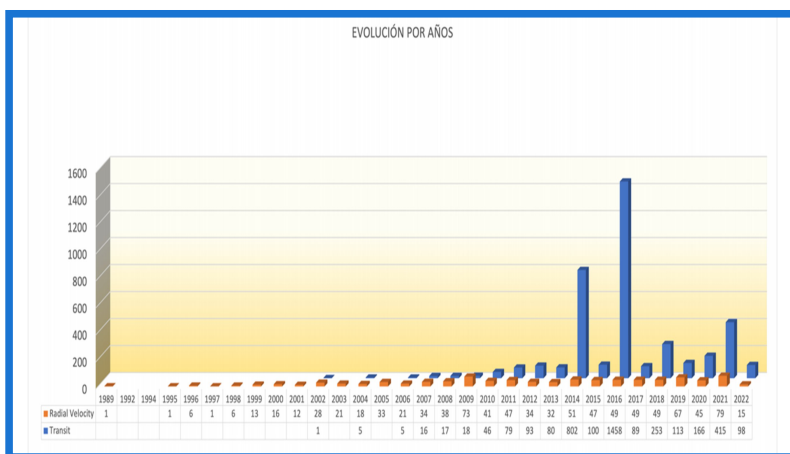


Fig. 1.3 – comparativa del número de exoplanetas descubiertos por los métodos de tránsito y velocidad radial. Mientras que el método velocidad radial mantiene un ritmo suave y constante, el método de tránsito crece de manera significativa, con picos, a partir del análisis de los datos recogidos por los telescopios espaciales.

Segundo ejercicio: Distribución espacial de exoplanetas.

Vamos a intentar ahora representaciones gráficas de la ubicación espacial de los exoplanetas conocidos y confirmados. Para ello copiaremos a una pestaña nueva, a partir del fichero original importado, el contenido de las siguientes columnas: 'pl_name', 'discoverymethod', 'ra', 'dec', 'sy_dist' {ra, dec, sy_dist}, que son respectivamente la posición en coordenadas ecuatoriales (ascensión recta y declinación, en formato decimal), y la distancia (en parsecs) de la estrella anfitriona, por tanto, del sistema planetario asociado a la misma.

Si observamos con cierto cuidado el contenido de la columna 'pl_name' (nombre asignado al exoplaneta), podemos ver que todos los registros acaban en una letra minúscula, (b, c, d, ...), precedida de otros caracteres. Esta última letra es la que identifica al planeta dentro del sistema planetario al que pertenece, y al que se le asigna el nombre de la estrella anfitriona. La notación sigue el orden de su descubrimiento (b, para el primero de los planetas del sistema; c, para el segundo; etc.). En cuanto al nombre de la estrella, indica el del catálogo que la contiene (HD, 2MASS, o alfa Tauri, beta UMi, ...) y presenta toda una casuística propia, incluidos los que asumen el nombre de la misión de exploración: Kepler, K2, CoRoT, WASP...). Los relativos a la misión TESS pueden identificarse en algunos casos por TOI (Tess Object Interest), haciendo alusión a la fase de análisis previa a la confirmación.

Por lo relevante de su número, vamos a asociar un indicador específico para los casos

'Kepler' y 'K2'. Para ello añadiremos una columna a nuestra tabla cuyo contenido sea 'Kepler' para los primeros y 'K2' para los segundos. Al resto les pondremos el indicador 'Otros'. Esto puede hacerse con facilidad utilizando una función sencilla del propio Excel aplicada a esta columna añadida y asumiendo que la columna A es 'pl_name':

```
=SI(EXTRAE(A2;1;2)="K2";"K2";SI(EXTRAE(A2;1;6)="Kepler";"Kepler";"Otros"))
```

a) Representación espacial de exoplanetas. Después de ordenar la tabla por la columna 'discoverymethod' es fácil obtener el siguiente gráfico:

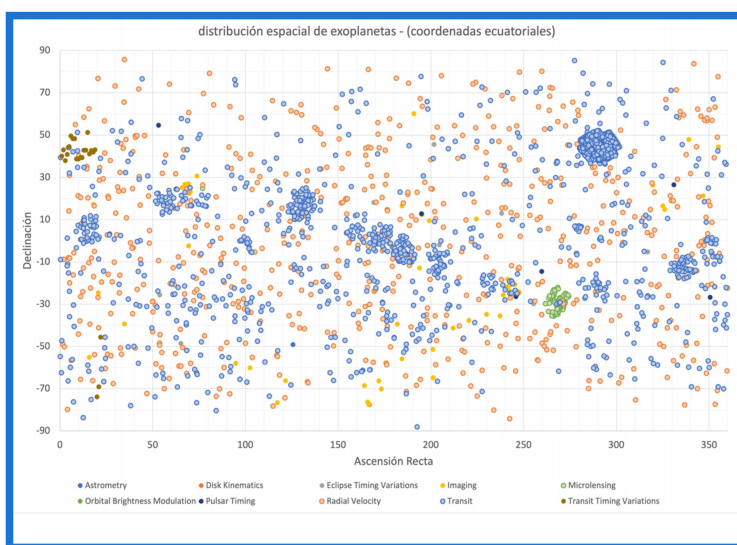


Fig. 1.4 – distribución de exoplanetas en coordenadas ecuatoriales. Se representan series diferenciadas por el método de detección. Como ya sabemos del ejercicio anterior, las dos series más pobladas son las correspondientes a los métodos de 'tránsito' y 'velocidad radial' respectivamente.

A simple vista se puede concluir que no hay una dirección privilegiada en el cielo que acumule exoplanetas, o dicho de otra manera, que carezca de ellos. Aunque, también a simple vista, se aprecian ciertas concentraciones que parecen significativas y que conviene analizar más despacio.

También puede ser interesante superponer al gráfico de la fig. 1.4 la posición de la galaxia en este mapa de coordenadas ecuatoriales. Véase la fig. 1 en el anexo.

b) Si nos centramos en las concentraciones de planetas que son perceptibles en la fig. 1.4, vemos que básicamente corresponden a exoplanetas descubiertos por el método del tránsito, y en concreto por la misión Kepler y Kepler en su fase extendida K2.

Vamos a representar sólo el mapa de estos planetas (fig. 1.5). Para ello convendrá ordenar la tabla por la columna añadida anteriormente con los indicadores {Kepler, K2, Otros}, que facilitará la construcción de la gráfica:

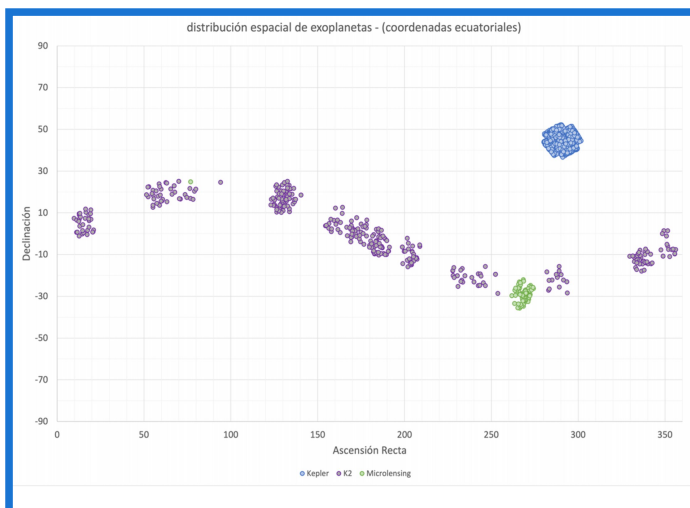


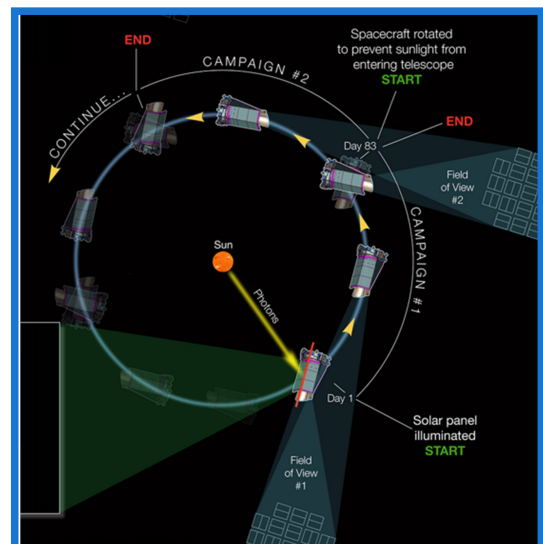
Fig. 1.5 – distribución espacial (en coordenadas ecuatoriales) de los exoplanetas descubiertos por la misión Kepler (azul), Kepler fase K2 (morado). En verde se sitúan también exoplanetas detectados por el método 'microlente' (que no tienen relación con la misión Kepler).

Indudablemente esta distribución contrasta con la aparente homogeneidad espacial concluida a partir de la fig. 1.4 lo que hace razonable indagar algo más acerca de las características propias de las observaciones que han dado lugar a estos datos.

La misión Kepler puso en órbita este telescopio espacial en marzo de 2009 con una duración prevista de 3.5 años. Su objetivo era inicialmente observar de forma ininterm-

pida y continua una misma región del cielo próxima al plano de la galaxia y centrada en torno a los 290° AR y 45° DEC, observando simultáneamente unas 150.000 estrellas. Recuérdese que el método del tránsito requiere una detección de al menos tres tránsitos que permitan validar un descubrimiento positivo e identificado, lo que justifica esta observación continuada de unas mismas estrellas-objetivo. Esto es lo que explica la concentración de exoplanetas detectados entre los 280° - 305° AR y 38° - 53° DEC.

A finales de 2012 se vio posible ampliar la misión otros cuatro años, pero en 2013 se estropeó uno de los giróscopos utilizados para la correcta orientación del telescopio. Era el segundo que dejaba de funcionar, (el primero lo hizo al poco tiempo de su lanzamiento). Durante algunos meses la NASA intentó la reparación de al menos uno de los dos giróscopos estropeados ya que de los cuatro con que Kepler contaba son necesarios tres para mantenerlo orientado, pero no fue posible y la misión entró en la fase extendida llamada K2. En esta etapa se utilizaron los paneles solares de alimentación del telescopio como sensores de orientación, lo que obligaba a dar 'un salto' en la dirección de observación cada tres meses, (véase la figura adjunta). Esta circunstancia explica



la distribución de exoplanetas K2 de la fig. 1.5. A finales de 2016 se dio por concluida la misión Kepler/K2.

Todavía nos queda una sorpresa más. Si hacemos zoom sobre la región ‘Kepler’, el detalle fino de la distribución de planetas en la misma es el que se muestra en la fig. 1.6. Indudablemente el espaciado geométrico debe tener alguna razón también instrumental. Compárese con la imagen del sensor instalado en el telescopio Kepler y quedará todo aclarado.

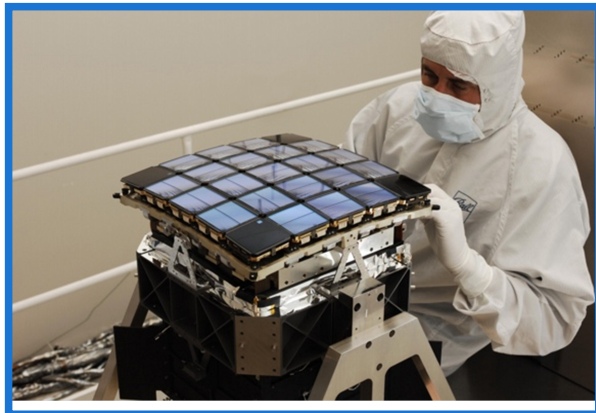
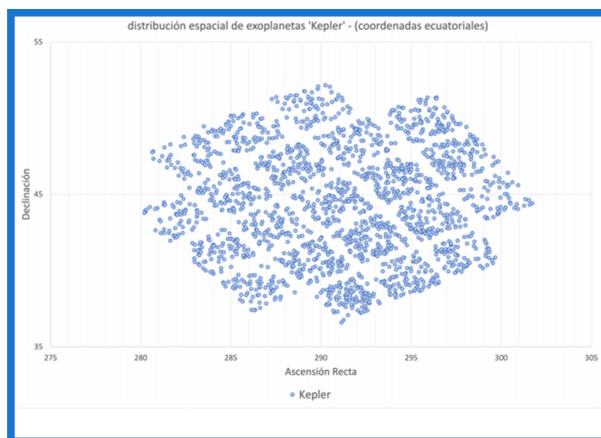


Fig. 1.6 – ampliación de la región Kepler

c) Sólo un comentario adicional sobre otro grupo llamativo y representado en la fig. 5 en color verde. Corresponde al conjunto de exoplanetas detectados por el [método de ‘microlente’ gravitacional](#). Ya no nos extraña

encontrar agrupamientos y sabemos que esto no significa que los planetas se encuentren en sitios privilegiados de la galaxia por alguna razón. Simplemente las técnicas de detección, y su corta historia, obligan a centrar las observaciones en posiciones estables durante periodos de tiempo más o menos amplios. En este caso sobre el brazo de Sagitario en dirección al centro galáctico.

Si volvemos a la tabla de datos, veremos que todos los exoplanetas catalogados por este método tienen nombres del tipo ‘KMT’, ‘MOA’ y ‘OGLE’. Corresponden respectivamente a los proyectos: ‘[The Korea Microlensing Telescope Network \(KMTNet\)](#)’, ‘[Microlensing Observations in Astrophysics \(MOA\)](#)’ y ‘[The Optical Gravitational Lensing Experiment \(OGLE\)](#)’, con observatorios en el hemisferio sur y centrados los tres en esa misma región del cielo con más de 100 millones de estrellas.

Tercer ejercicio: Distribución de exoplanetas por distancia desde la Tierra

Para terminar esta primera toma de contacto con los datos sobre exoplanetas puede ser interesante fijarnos en la distancia a que se encuentran. Esto nos dará una idea de la sensibilidad de la instrumentación y técnicas disponibles por ahora y su capacidad de profundidad en el cielo para detectarlos.

a) En una primera aproximación planteamos un esquema simple donde se representan los exoplanetas con su respectiva distancia a nuestra posición. Esta distancia se encuentra en la columna ‘*sy_dist*’ que, como ya se ha comentado, viene expresada en parsec. Aprovechamos para diferenciar por método de detección estableciendo series sólo para los más significativos: tránsito, velocidad radial, microlentes gravitatorias e ‘imagen’,

siendo esta última relevante en tanto que supone el único método de detección directa.

El gráfico planteado se muestra en la fig. 1.7. En el eje horizontal aparece el número de exoplanetas representado para cada una de las series, siendo éste un dato poco relevante aquí ya que ha quedado mejor explicado en las fig. 1.1, 1.2 y 1.3. En el eje de ordenadas se representan las distancias en parsec.

Como se puede comprobar, el método ‘microlensing’ es significativamente más sensible que el resto, presentando una capacidad de profundizar en el cielo hasta unos 8.000 pc (próximo al centro galáctico). Por el contrario, el método directo ‘imaging’ es el más limitado y los planetas detectados no superan nunca los 200 pc. También el método de velocidad radial es menos profundo que el de tránsitos.

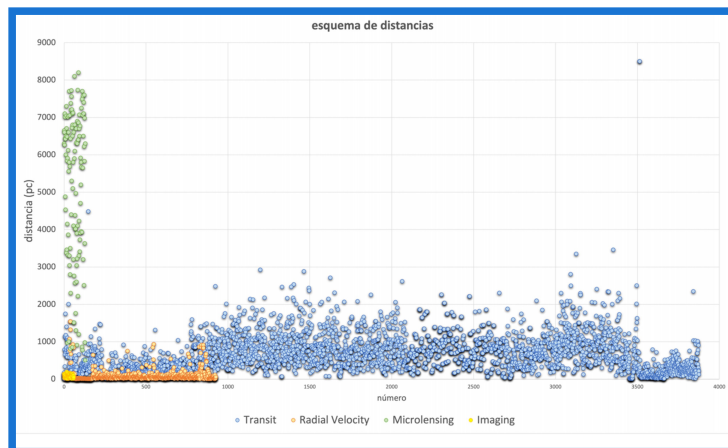


Fig. 1.7 – esquema de distribución de distancias según los diferentes métodos de detección. El par de planetas por encima de 8.000 pc en el método de tránsito corresponden a dos sistemas en las estrellas SWEEPS-11 y SWEEPS-4 (ver tabla en el fichero de datos). Evidentemente son datos a contemplar con cierta reserva puesto que, como puede verse en dicha tabla, corresponden a sendos registros con ausencia significativa de otros parámetros relativos a estas estrellas, incluidos los márgenes de error de esta propia medida.

Si hacemos zoom sobre el vértice ‘origen’ en el gráfico anterior tendremos algo más de detalle en el rango más cercano.

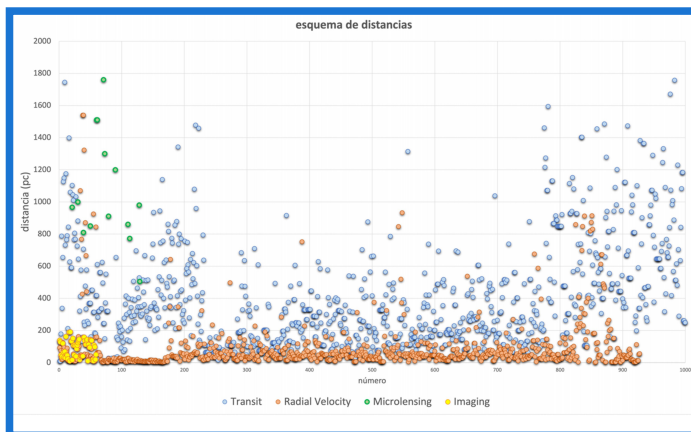


Fig. 1.8 – detalle en el origen de coordenadas de la fig. 1.7

No obstante, resulta evidente que podemos intentar un gráfico más significativo sobre las distancias y dirección en el cielo (ra , dec) que tenemos disponible en este fichero. Pero para ello hay que convertir el vector (ra - dec , sy_disc) a coordenadas cartesianas. Añadiremos tres columnas a la última tabla que estamos manejando: { st_x , st_y , st_z }, que corresponden respectivamente a los ejes (x , y , z) para la representación en un diagrama cartesiano de tres dimensiones.

El contenido de estas columnas es el que corresponde a la transformación de coordenadas:

$$\begin{aligned}x &= distancia \cdot \cos(declinación) \cdot \cos(ascensión\ recta) \\ y &= distancia \cdot \cos(declinación) \cdot \sin(ascensión\ recta) \\ z &= distancia \cdot \coseno(declinación)\end{aligned}$$

Donde los ángulos ($ascensión\ recta$, $declinación$) deben expresarse en radianes y las distancias (en parsec) mantendrán esta unidad en las nuevas coordenadas. Como en algunos registros el valor del parámetro ‘ sy_dist ’ no está informado, tendremos en cuenta esta circunstancia para evitar errores en las fórmulas de transformación que finalmente pueden ser de la forma:

$$\begin{aligned}st_x &= SI(Fn < > "" ; Fn * SENO(RADIANES(En)) * COS(RADIANES(Dn)); "") \\ st_y &= SI(Fn < > "" ; Fn * SENO(RADIANES(En)) * SENO(RADIANES(Dn)); "") \\ st_z &= SI(Fn < > "" ; Fn * COS(RADIANES(En)); "")\end{aligned}$$

siempre que en nuestra tabla la distancia se encuentre en la columna F y *ascensión recta* y *declinación* en las columnas D y E respectivamente. Si no fuera así, se sustituyen las referencias a estas columnas por su valor correcto. El indicador 'n' corresponde a la fila, e indica que las expresiones anteriores deben extenderse a todas las filas del fichero. Basta con hacerlo para la primera, n=2, y luego copiar/pegar estas fórmulas para el resto.

Ahora ya estamos en condiciones de poder hacer una representación cartesiana en 3D de los exoplanetas contenidos en nuestro fichero:

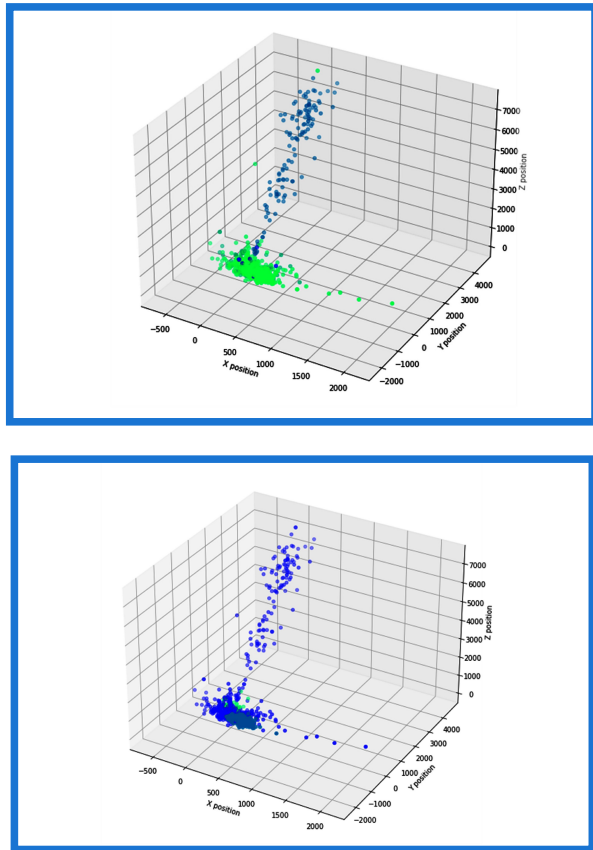


Fig. 1.9 – representación cartesiana de distribución de exoplanetas y distancias desde la Tierra.

En la parte superior de la fig. 1.9 se representan los exoplanetas diferenciados por el método de detección y resultan más destacados los métodos de microlente (azul

claro) y tránsito (verde). El resto queda enmascarado en la zona de distancias más cortas. En la parte inferior de la figura 1.9 se discrimina por lo que hemos llamado 'origen/misión' = {Kepler: azul verde, K2: azul claro, Otros: azul oscuro}. Igualmente la zona próxima al origen de coordenadas {0, 0, 0} resulta confusa si no se hace zoom en la misma, debido a la mayor profundidad de los 'chorros' microlente y Kepler.

Para la representación tridimensional de las figs. 1.9 Excel puede resultar insuficiente si no se le añade alguna extensión de análisis de datos (del tipo XLSTAT). En este caso se ha utilizado Python, (en el anexo se dan detalles).

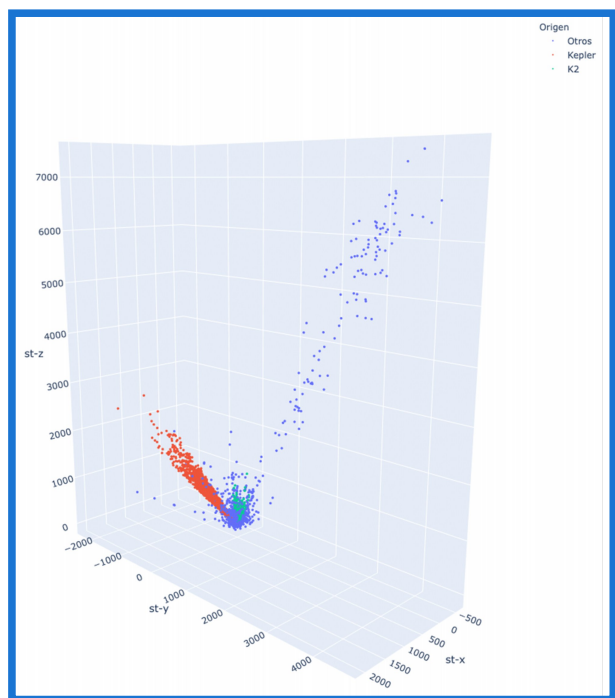


Fig. 1.10 – ampliación de la fig. 1.9 con otra perspectiva y un mapa de colores diferente.

La conclusión final es que la existencia de sistemas planetarios es algo generalizado, como puede deducirse de la distribución espacial que muestra la fig. 1.4. Pero el catálogo de exoplanetas conocidos hasta el momento está muy condicionado por la corta

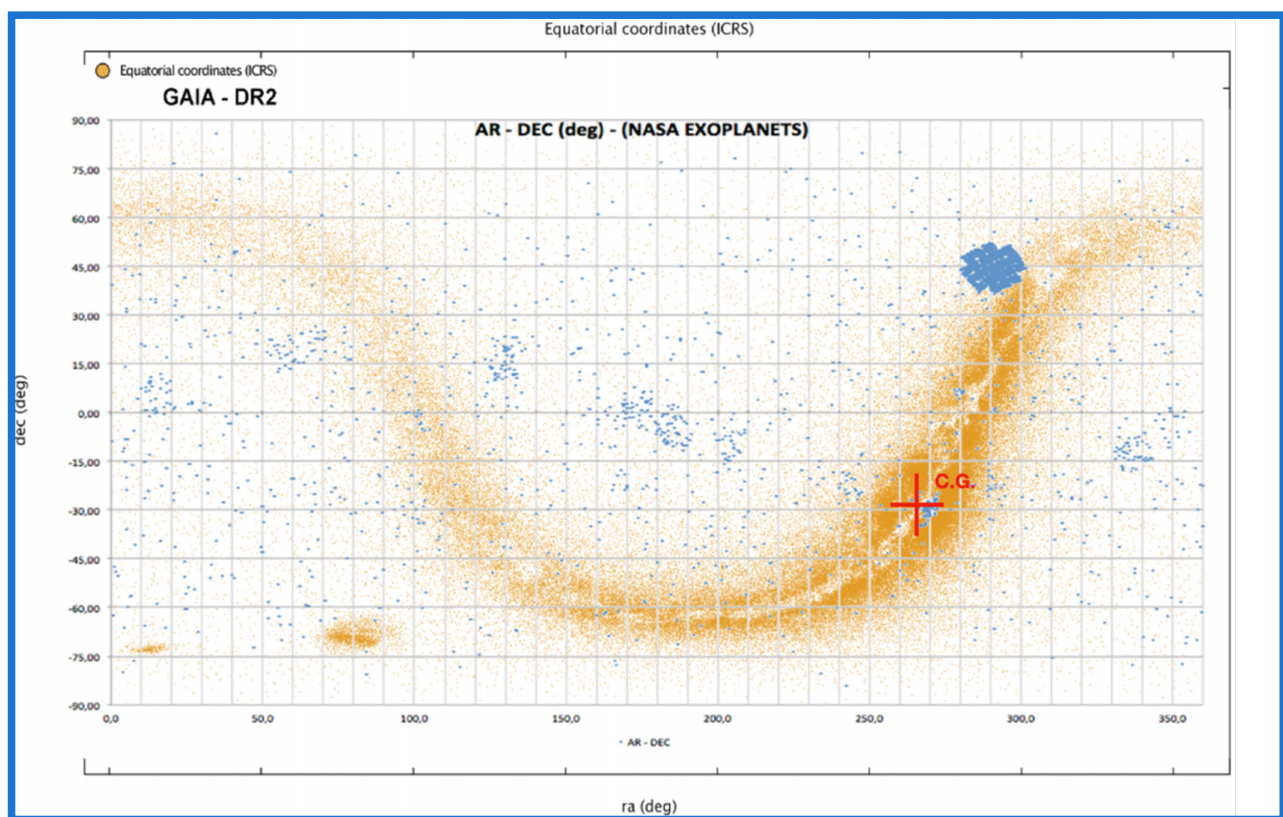
historia de las técnicas de detección y, especialmente, por el tiempo dedicado a la exploración de direcciones concretas del cielo, como muestran los dos chorros de las figs. 1.9 y 1.10.

Terminamos aquí este primer artículo. Se aconseja realizar los ejercicios/gráficos,

propuestos a modo de ejemplo, u otros similares. Habrá continuidad en sucesivas publicaciones como se ha indicado en la introducción. Si alguno de los ejercicios presenta alguna duda, o para cualquier otra cuestión relacionada, podéis contactar conmigo: jalvaro@citelan.es

ANEXO 1

a) Superposición de la Vía Láctea sobre la fig. 1.4 simplificada



La imagen de la galaxia se ha generado utilizando las estrellas de magnitud < 16 a partir del catálogo [Gaia-DR2](#). Se ha marcado el centro galáctico (C.G.).

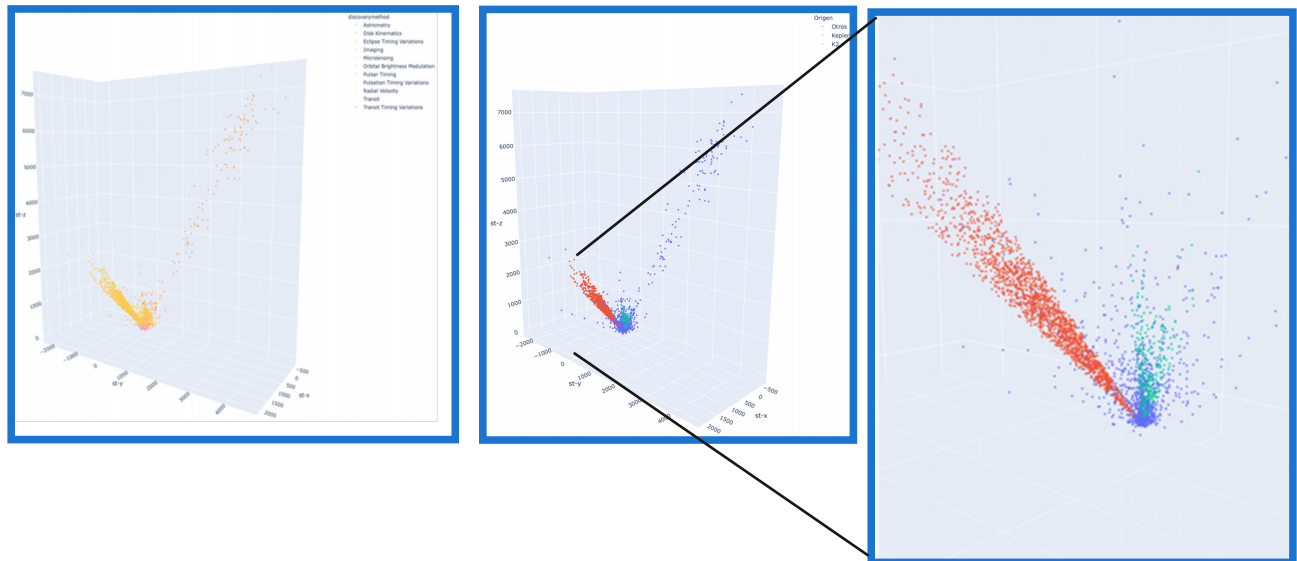
b) Representación tridimensional de la distribución espacial de exoplanetas en coordenadas cartesianas.

Para las figuras 1.9 y 1.10 se ha utilizado código en Python. En concreto se ha optado por una solución que da como salida un gráfico visualizable como un fichero .html con capacidad para ser rotado en cualquiera de los ejes y sobre el que se puede hacer zoom.

```
# librerías utilizadas
import numpy as np
import pandas as pd
import plotly
import plotly.express as px

# lectura del fichero exportado desde excel (formato .csv), delimitador ','.
df = pd.read_csv('exoplanets-1-mapa3d.csv', sep=',',
                 names=['pl-name', 'Origen', 'discoverymethod', 'ra', 'dec', 'sy_dist',
                       'st-x', 'st-y', 'st-z'], skiprows=1)
df.head()

# generación del gráfico en formato 3D y exportación a formato .html
fig = px.scatter_3d(df, x='st-x', y='st-y', z='st-z', color='discoverymethod')
fig.update_traces(marker=dict(size=2))
plotly.offline.plot(fig, filename="mapa_3d.html")
```



2ª PARTE



En esta segunda parte sobre el trabajo con datos reales a partir del fichero “[Planetary Systems Composite Data](#)”^[1] entraremos en un mayor nivel de detalle con ejercicios que pretenden acercarnos a un mejor conocimiento de lo que una base de datos de estas características, estadísticamente ya muy significativa, puede ofrecernos: los sistemas planetarios, la zona de habitabilidad y el tipo de estrellas.

El objetivo esta vez es analizar algunas correlaciones de interés entre los parámetros que integran el fichero. Puede resultar útil empezar por tener

una visión general de las distribuciones relacionadas, en principio siempre por pares. Véase la figura de la página siguiente.

[1] "This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program."

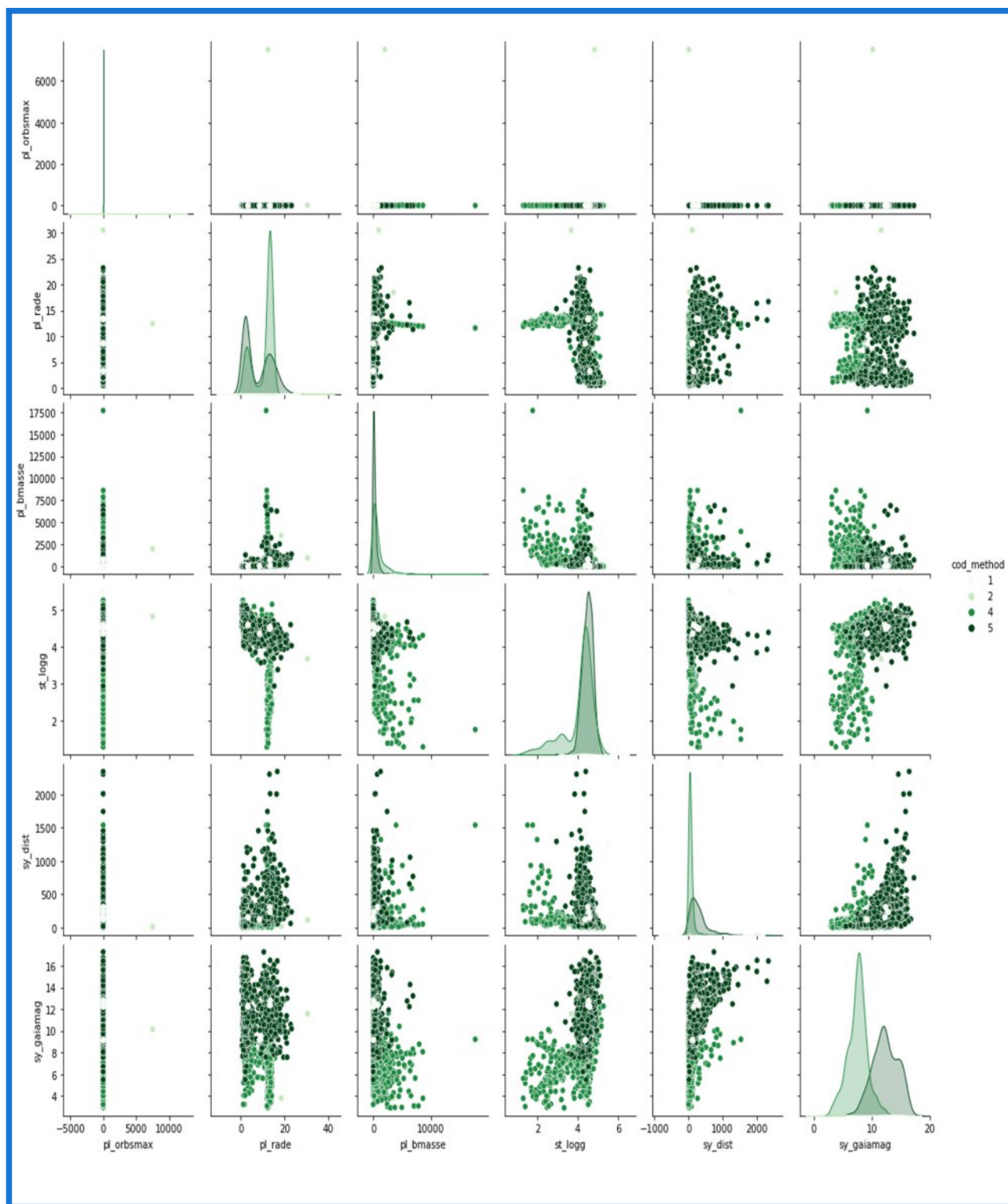


Fig. 2.1 – Múltiples correlaciones por pares de variables

No vamos a entrar en mucho detalle, por ahora, sobre lo que nos muestra este gráfico desarrollado con un poco de código Python, (ver anexo para más explicaciones). Lo que se pretende aquí es centrarse en los parámetros

a priori más significativos, tanto de los planetas como de las estrellas anfitrionas, y relacionarlos entre si (de dos en dos) con objeto de ver si en un primer análisis se aprecian dependencias significativas a simple

vista. De ser así, éstas serían las candidatas a considerar con mayor atención en un análisis más detallado, que ya volveremos a hacer con Excel en los ejercicios que siguen.

En primer lugar se ha simplificado la clasificación de valores del fichero en el método de detección *'discoverymethod'*, asignando los valores de {2=*'imaging'*, 3=*'microlensing'*, 4=*'radial velocity'*, 5=*'transit'* y 1=*'resto'*} en un campo nuevo *'cod_method'* con objeto de hacer más sencillo este posible análisis, (algo que ya se hizo en el artículo anterior "Exoplanetas-1: fig. 7 y 8"), y lo primero que se observa es que, en la fig. 2.1, no aparece representado el color correspondiente a *'3=microlensing'*, lo que está indicando algún tipo de problema con los datos relativos a este método de detección que habrá que estudiar.

Por lo demás, y sin entrar en mucho detalle como ya se ha dicho, llaman la atención algunas distribuciones que pueden significar dependencias directas entre las variables relacionadas y que será el objeto de todo el trabajo por hacer. Por ejemplo, son llamativos los histogramas (diagonal en fig. 2.1) que ponen de manifiesto alguna dispersión, (con diferencias incluso según el método de detección), como los relativos a *'sy_gaiamag'* (magnitud fotométrica aparente de la estrella en la banda espectral de Gaia), que debería estar muy directamente relacionada con *'sy_dist'* (distancia), y que muestra dos poblaciones diferentes para 4=*velocidad radial* y 5=*tránsito*, como se puede apreciar en los gráficos (*sy_gaiamag*, *sy_dist*), algo que todavía es más notorio en (*sy_gaiamag*, *st_logg* ^[2]), o en (*st_logg*, *sy_dist*). Y también el histograma de *'pl_rade'*^[3] es destacable. Pero esto lo iremos viendo en detalle.

Sistemas planetarios más allá del Sistema Solar.

Llegados aquí ya sabemos que no hay direcciones privilegiadas en el cielo donde aparezcan los exoplanetas, sino que en general los podremos encontrar en la mayoría de las estrellas observadas. Si el análisis preliminar hecho en el artículo anterior mostraba sesgos en este sentido, debió quedar claro que las razones al respecto estaban tanto en la corta historia de las técnicas empleadas como en las limitaciones propias de las mismas, así como en el tiempo, o paciencia, dedicado a mantener la constancia de las observaciones.

Ahora nos podemos plantear si los sistemas planetarios extrasolares conocidos, a pesar de ser todavía una cantidad ínfima (meramente testimonial respecto a los miles de millones que debe haber), se parecen a nuestro Sistema Solar.

En el fichero que estamos trabajando hay dos parámetros que podemos rastrear sin mucho esfuerzo y sin necesidad de mayor análisis. Son *'sy_snum'* y *'sy_pnum'*. El primero de ellos indica el número de estrellas que hay en el sistema estelar contemplado y, como se puede comprobar, en algunos casos se trata de estrellas dobles e incluso un conjunto de 3 o hasta 4. En cuanto a *'sy_pnum'*, éste indica el número de exoplanetas conocidos en estos sistemas planetarios. Como también se puede comprobar, en algunos se conocen hasta 6 o incluso 8, pero en general sólo son 1 o 2 los conocidos.

Vamos a comparar los sistemas planetarios con el nuestro, estudiando dos aspectos:

[2] *st_logg* = Stellar Surface Gravity [$\log_{10}(\text{cm/s}^2)$]. Ver el repositorio de los datos.

[3] *pl_rade* = radio del planeta en unidades del radio terrestre. El otro radio contemplado en el fichero, *pl_radj*, es el mismo radio expresado en unidades del radio de Júpiter. Lo mismo ocurre con las masas.

a) ¿Cómo de lejos se encuentran estos exoplanetas de sus respectivas estrellas?

Para indagar esta cuestión empezaremos por copiar a una nueva pestaña algunas de las columnas del fichero, con objeto de hacer menos pesados los diferentes procesos de ordenación y manipulación necesarios. Vamos a utilizar los siguientes parámetros^[4]:

pl_name

discoverymethod

pl_orbper (periodo orbital – en días)

pl_orbsmax (semieje eje mayor de la órbita – en UA, unidades astronómicas)

pl_rade (radio del planeta en unidades terrestres)

pl_bmasse (masa del planeta en unidades terrestres)

st_spectype (tipo espectral de la estrella)

st_teff (temperatura superficial efectiva – en grados Kelvin)

st_lum (logaritmo de la luminosidad – en unidades solares: Sol = 1)

st_logg (gravedad superficial de la estrella – $\log_{10}(\text{cm/s}^2)$)

st_age (edad de la estrella – en Gyr: 10^9 años)

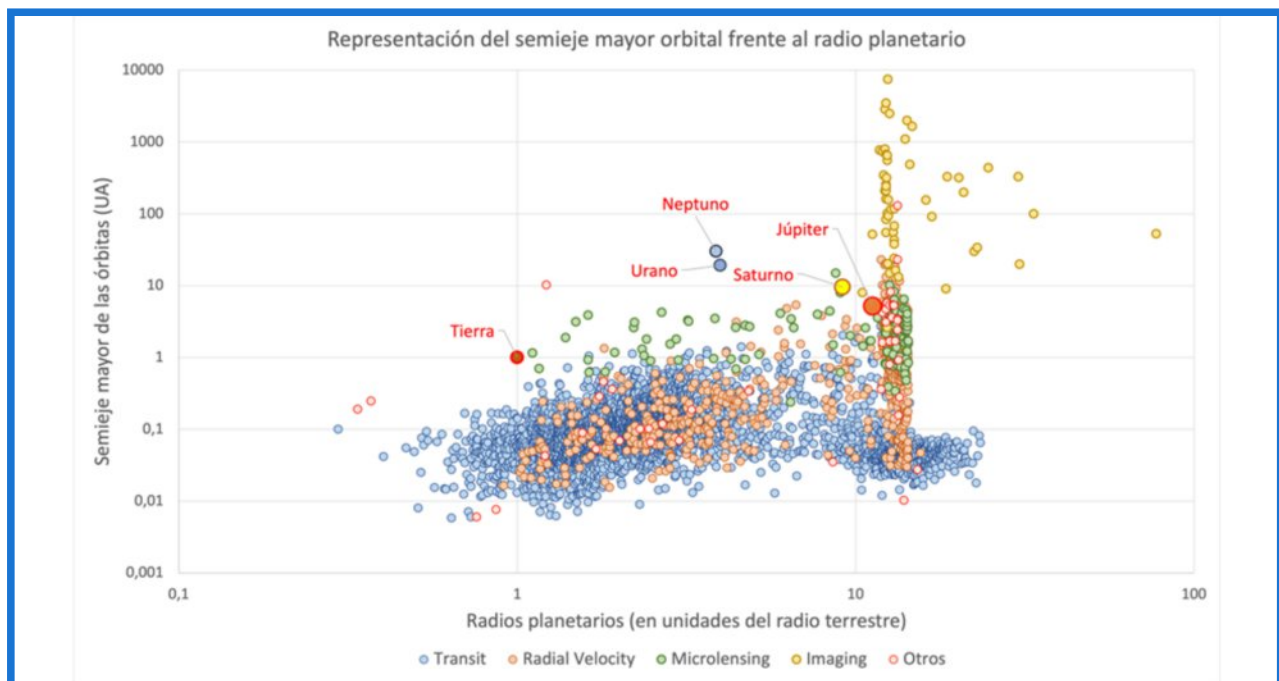
sy_gaiamag (magnitud fotométrica según Gaia)

- Para simplificar los gráficos añadiremos la columna '*cod_method*', (antes ya contemplada), en la que se agrupan algunos de los métodos de detección menos significativos. Su contenido se reduce a:

cod_method={'Imaging','microlensing','Radial Velocity','Transit','Otros'}

- Ordenaremos esta tabla por *cod_method* y *pl_rade*, esto es, por el método de detección y el radio del planeta, con objeto de representar la distribución de distancias

Fig. 2.2 – Representación del semieje mayor orbital en función del radio planetario diferenciando series por el método de detección. Los ejes están representados en escala logarítmica. Se han añadido las posiciones en el gráfico de la Tierra, Júpiter, Saturno, Urano y Neptuno, como referencias comparativas propias del Sistema Solar.



[4] Si alguno de estos parámetros no se tiene porque no se descargó inicialmente del fichero, puede ignorarse, pero se recomienda en cualquier caso repetir el proceso de descarga incluyendo las columnas que se echen en falta.

[5] Debe hacerse notar que estamos asumiendo el semieje mayor orbital como la distancia planeta-estrella. Sería más riguroso plantearse una distancia media calculada a partir de los parámetros semieje mayor y excentricidad de la órbita, pero aquí esta aproximación es aceptable.

planeta-estrella^[5] frente al radio del planeta con diferentes series según el método de detección. El resultado se muestra en la fig. 2.2.

Lo primero a destacar en la fig. 2.2 es que la mayoría de los exoplanetas conocidos quedan por debajo de la línea de trazos que marca 1 UA, es decir, si estos estuvieran en el Sistema Solar, serían todos planetas interiores a la Tierra, como Venus y Mercurio. Esta mayoría es concretamente de un 86.6%, (como puede calcularse fácilmente a partir de los datos que venimos manejando ^[6]).

¿Significa esto que ésta es una característica propia de los sistemas planetarios extrasolares? Evidentemente no. Lo que se pone de relieve aquí es que las técnicas actuales tienen más fácil la detección de planetas próximos a sus respectivas estrellas. Esto es especialmente manifiesto en el caso de la detección por el método del tránsito, incluso para planetas grandes, (radios $>10 R_T$), ya que se requieren al menos tres tránsitos para confirmar la existencia identificada de un exoplaneta. Si la distancia del planeta a su

estrella es grande, también lo será su periodo orbital, y no ha habido tiempo suficiente aún para que el número de exoplanetas conocidos alejados sea mayor por esta razón.

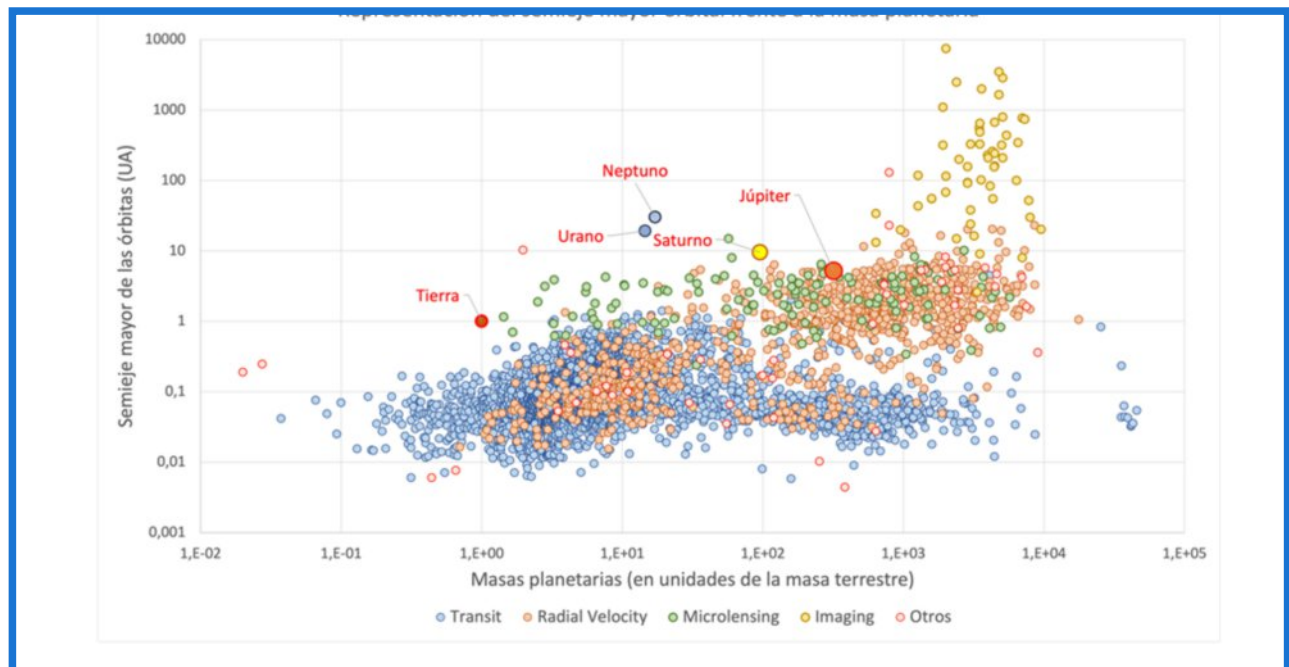
En el caso de la detección por el método directo de Imagen por el contrario, resulta favorable la condición de que sí haya una mayor lejanía del planeta a su estrella, donde además también debe ayudar el que se trate de planetas de gran radio. Esto se ve claramente en la fig. 2.2.

Si planteamos una representación similar a la anterior pero sustituyendo el radio del planeta por su masa, lo que tenemos es lo que nos muestra la fig. 2.3.

También aquí las escalas de los ejes son logarítmicas para dar más detalle al gráfico. (Véase en la fig. 2.1 que los bloques de la parrilla que afectan al semieje mayor orbital -fila 1 y columna 1- no son representativos por la poca dispersión de los datos frente a algunos valores muy extremos que incluso pueden ser erróneos).

El gráfico 2.3 matiza un poco las

Fig. 2.3 – Representación del semieje mayor orbital en función de las masas.



[6] Debe tenerse en cuenta que hay 265 exoplanetas de los que no se tiene información sobre su semieje mayor orbital y que han sido excluidos del cálculo del porcentaje

conclusiones anteriores al poner de manifiesto igualmente que el método de velocidad radial es más sensible a planetas masivos, como era de esperar. Véase el amplio conjunto por este método con masas $> 100 M_T$. También resulta evidente la amplia mayoría por debajo de 1 UA.

Por último, haremos un simple ejercicio sobre la tercera Ley de Kepler:

$$\frac{T^2}{a^3} = Cte$$

donde, T es el periodo orbital y a el semieje mayor de la órbita.

Para ello representaremos el periodo orbital frente al semieje mayor. Véase la fig. 2.4, donde se puede observar una muy buena correlación entre estas dos variables ajustada a la ley de potencia esperada:

$$[1] \quad T = Cte. a^{3/2}$$

El exponente, en el ajuste calculado por Excel a partir de los datos con los que estamos trabajando, es 1.4664 que resulta compatible con la tercera ley de Kepler, según

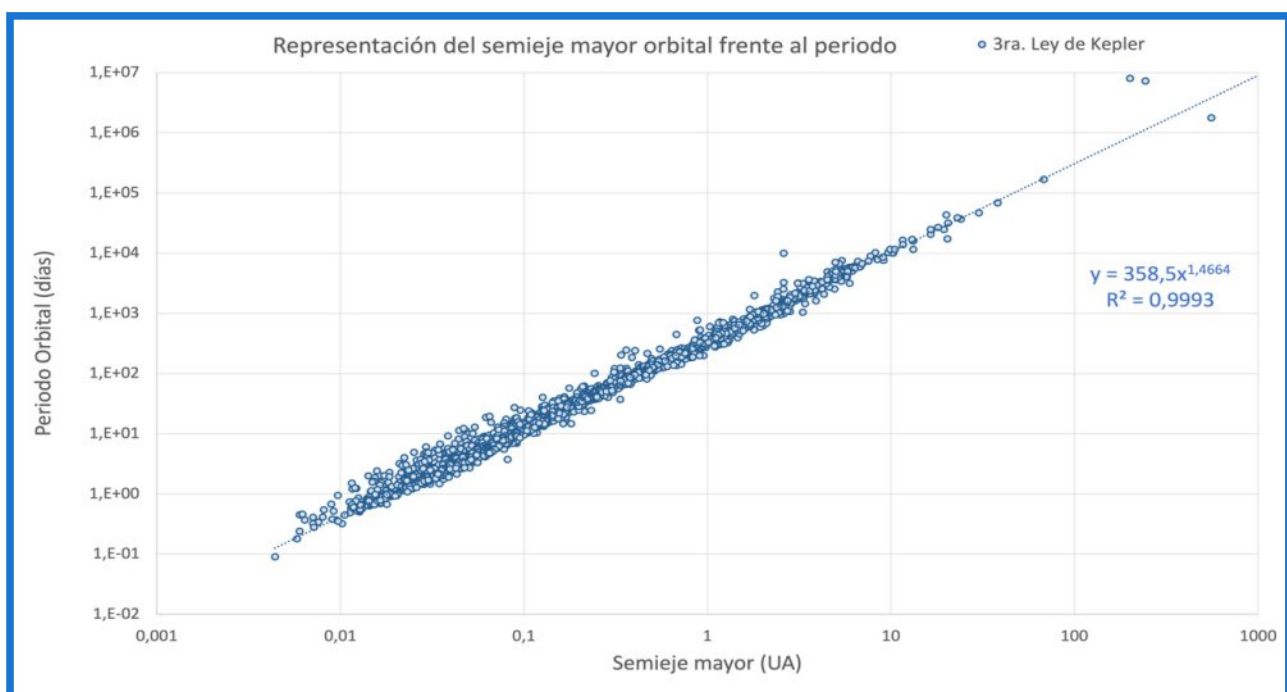
la cual este exponente debería ser 1.5 ($=3/2$), según [1], dada la incertidumbre en los datos reflejada en los valores de error (+/-) asociados a estos parámetros. Véase el anexo.

b) La zona de habitabilidad.

El segundo estudio será: ¿cuántos, (y cuáles), de los exoplanetas conocidos son potencialmente habitables? Ésta es una cuestión que puede resultar interesante y motivadora para trabajarla en el aula. Aunque ya sabemos que los planetas descubiertos hasta el momento están en su gran mayoría muy próximos a sus estrellas, (debido a la capacidad de las técnicas de detección disponibles), el que un planeta se encuentre dentro de la zona de habitabilidad depende no sólo de la distancia a la estrella, sino también de las características de ésta última.

Para este ejercicio utilizaremos el criterio de habitabilidad definido por el *PHL (Planetary Habitability Laboratory)* –[David R. Underwood (2003) & Franck Selsis (2007)]– que define dos radios, o distancias desde la estrella, entre los cuales se asumen condiciones de habitabilidad. Estos dos radios r_i y r_o son respectivamente los radios interior y exterior. Por debajo

Fig. 2.4 – Periodo orbital frente a semieje mayor de la órbita. Tercera ley de Kepler



del primero las altas condiciones de insolación del planeta lo hacen incapaz de albergar vida y por fuera del radio exterior, llamado 'línea de hielo', ocurre por el contrario que el frío es causante de las condiciones de no habitabilidad.

Estos radios dependen básicamente de la temperatura superficial efectiva de la estrella y de su luminosidad, y se definen así:

$$[2] \quad \begin{aligned} r_0 &= [r_{0s} - a_0(T_{eff} - T_{\odot}) - b_0(T_{eff} - T_{\odot})^2] \sqrt{L} \\ r_i &= [r_{is} - a_i(T_{eff} - T_{\odot}) - b_i(T_{eff} - T_{\odot})^2] \sqrt{L} \end{aligned}$$

Siendo T_{eff} y L la temperatura efectiva y la luminosidad respectivamente de la estrella, y $T_{\odot}$ la temperatura superficial efectiva del Sol (≈ 5.700 K), mientras que el resto de parámetros toman los siguientes valores:

$$\begin{aligned} a_0 &= 1.3786 \cdot 10^{-4}, b_0 = 1.4286 \cdot 10^{-9}, \\ a_i &= 2.7619 \cdot 10^{-5}, b_i = 3.8095 \cdot 10^{-9}, \\ r_{0s} &= 1.77, r_{is} = 0.72 \end{aligned}$$

Los radios $\{r_0, r_i\}$ así calculados están expresados en unidades astronómicas.

La luminosidad se expresa en unidades solares, $L_*/L_{\odot}$

Como se puede ver, hasta ahora sólo se han definido dos radios $\{r_0, r_i\}$ que dependen exclusivamente de cada estrella. Es preciso determinar si cada planeta en cuestión está o no dentro del margen determinado por esos radios. Para ello se define igualmente un parámetro normalizado, llamado HZD, de la siguiente manera:

$$HZD = \frac{2r - r_0 - r_i}{r_0 - r_i}$$

donde r es el radio de la órbita planetaria, es decir, su distancia a la estrella. Utilizaremos

aquí el valor del semieje mayor orbital (a) de nuestro fichero para cada uno de los planetas, $r=a$.

La zona de habitabilidad se define en cada caso para $HZD_p=(-1,+1)$, y, más estrictamente, hablaremos de 'zona de confort' si el valor está entre $HZD_p=(-0.5,+0.5)$.

Vista esta introducción teórica, ahora debemos preparar adecuadamente los datos necesarios para resolver el problema planteado con los planetas de nuestro fichero. Vamos a necesitar:

Temperatura efectiva de la estrella – st_teff [K]

Luminosidad de la estrella – st_lum [$\log(\text{Solar})$]

Semieje mayor de la órbita del planeta – $pl_orbsmax$ [UA]

Tanto st_teff como $pl_orbsmax$ están en las unidades correctas y no necesitan ninguna manipulación. En cuanto a la luminosidad estelar, como se ve, en el fichero la tenemos expresada como el log en base 10 del cociente $L_*/L_{\odot}$ por tanto sólo debemos hacer una transformación simple aquí para el parámetro L , ($L=10^{(st_lum)}$), y lo tendremos en unidades solares, que son las contempladas en [2]. Después añadiremos las columnas necesarias para el resto de parámetros: r_i , r_0 y HZD.

Una vez calculado el parámetro HZD_p para cada planeta^[7] ya podremos representar la distribución de exoplanetas según su característica de habitabilidad (Fig 2.5).

Como era de esperar, la mayoría de los exoplanetas están fuera de la zona estimada de habitabilidad ($a < r_i$). Esto no significa que ésta sea la tónica dominante, al menos en esta proporción, sino que como ya sabemos las técnicas de detección actuales (en especial la más utilizada del tránsito) tiene más dificultades para la detección de planetas alejados de su estrella anfitriona.

[7] Hay que tener en cuenta que puede haber exoplanetas que no tengan información en alguna de las columnas contempladas: $pl_orbsmax$, st_teff , st_lum . Conviene, por tanto, emplear la función $=SI(condición;valor SI;valor NO)$ de Excel para dejar con valor nulo la columna HZD en cualquiera de estos casos y evitar el error correspondiente en los cálculos.

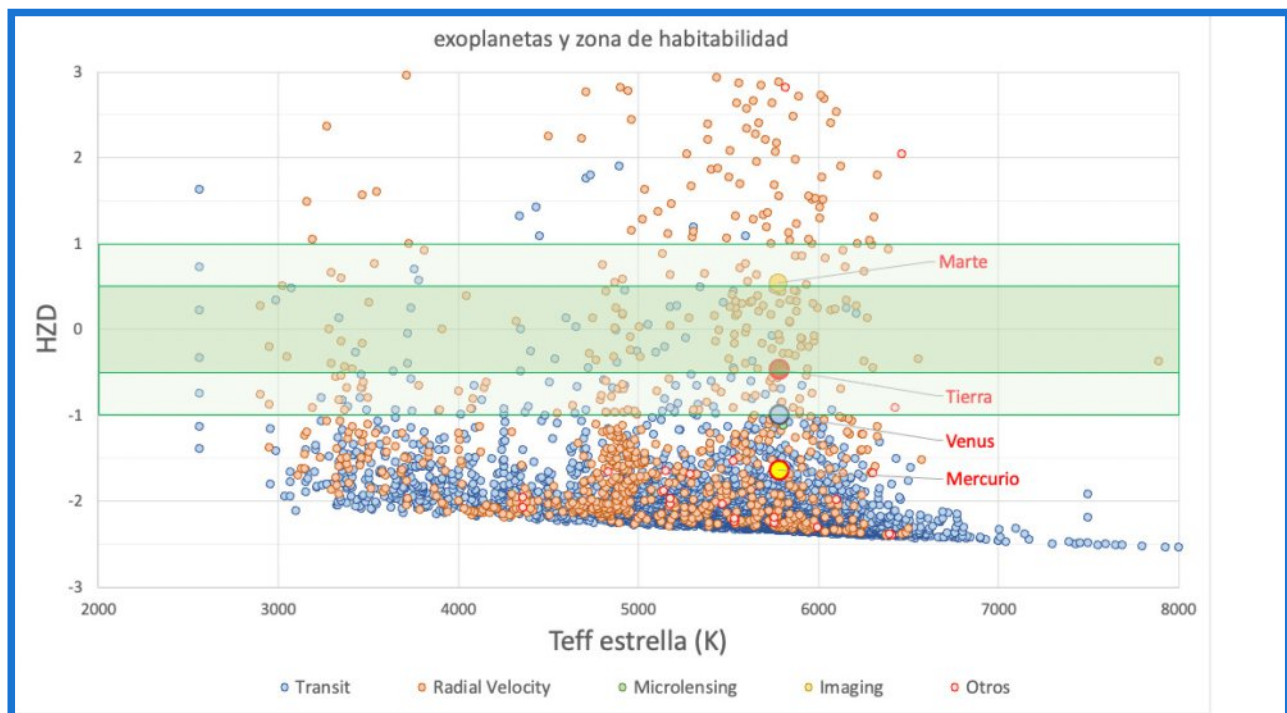


Fig. 2.5 – distribución de exoplanetas confirmados en función de su parámetro HZD.

En la fig. 2.5 se han remarcado las zonas de habitabilidad $HZD = (-1, +1)$ y de *confort* $(-0.5, +0.5)$. Se han añadido también, como referencias, las posiciones de los planetas Mercurio, Venus, la Tierra y Marte.

Quedan fuera del gráfico los planetas detectados por el método directo de '*imaging*' que, como también sabemos ya, son más fáciles de detectar si se encuentran a una mayor distancia de su estrella, todos ellos tienen un valor $HZD > 3$, y sólo uno de los detectados por el método '*microlensing*' tiene informados los datos de luminosidad y temperatura efectiva correspondientes a su estrella, lo que explica que este grupo no aparezca en la fig. 2.1.

c) Tipo de estrella

Para terminar esta segunda entrega haremos una exploración sencilla acerca de los tipos de estrellas representados en este catálogo. ¿Todos los tipos de estrellas tienen sistemas planetarios, o hay algún tipo que por sus características propias no lo haga posible?

Para ello revisaremos los tipos espectrales y también construiremos el diagrama H-R de

las estrellas presentes en este fichero. Es un buen momento para repasar en el aula qué es un diagrama Hertzsprung-Russell y las etapas evolutivas en la vida de las estrellas.

Para construir el diagrama H-R con el conjunto de estrellas que tenemos en este fichero basta representar su distribución en función de la temperatura efectiva y la luminosidad. Emplearemos los parámetros de las columnas st_teff y st_lum (en unidades solares, $L = 10^{(st_lum)}$, como en el ejercicio anterior). El resultado está en la fig. 2.6.

Como puede comprobarse, prácticamente todos los tipos estelares están bien representados, desde las enanas rojas a las gigantes rojas y también las azules, siendo especialmente numeroso el grupo de estrellas en la secuencia principal, (como era de esperar ya que las estrellas pasan la mayor parte de sus *vidas* en esta fase). Sin embargo, la rama de las estrellas masivas y muy calientes (gigantes azules) está menos poblado. Es cierto que estas estrellas tienen una vida más corta, pero sería interesante estudiar con más atención esta circunstancia para determinar si esta menor representación

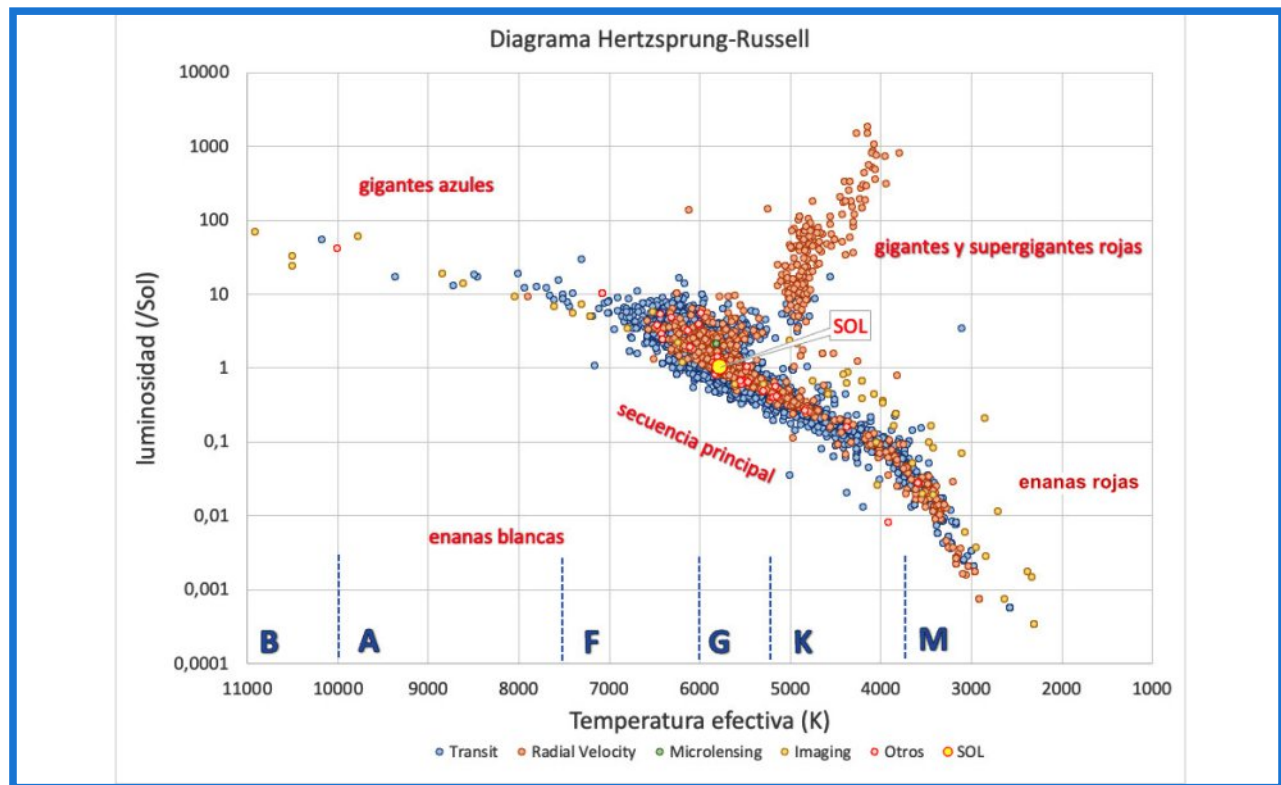


Fig. 2.6 – diagrama Hertzsprung-Russell de las estrellas anfitrionas de los exoplanetas confirmados. El eje de ordenadas se representa en escala logarítmica.

se debe a deficiencias de las propias técnicas de detección o si está indicando realmente una cierta escasez de sistemas planetarios en este tipo de estrellas.

Igualmente es significativa la ausencia de exoplanetas en estrellas del tipo *enanas blancas*, que no aparecen en el diagrama. También aquí habría que determinar si se trata de dificultades en la detección con las técnicas actuales, o bien está poniéndose de manifiesto que los sistemas planetarios no *sobreviven* a la fase explosiva de las gigantes rojas que deja como resto este tipo de estrellas.

En la fig. 2.6 están diferenciados, por colores, las series que venimos representando en función del método de detección, (se pueden extraer conclusiones al respecto igualmente), así como también están marcados los límites aproximados de los tipos espectrales, (M, K, G, F, A, B, O).

Hay más ejercicios que se pueden desarrollar aprovechando la información

contenida en el fichero '*Planetary System Composite Data*' que pueden resultar de gran interés didáctico. Por ejemplo, puede hablarse de magnitudes aparentes y magnitudes absolutas de las estrellas, y a partir de ello calcular la magnitud absoluta M de estas estrellas utilizando los valores contenidos en las columnas *sy_gaiamag* y *sy_dist* según:

$$M = m + 5 - 5 \log (d)$$

haciendo $m = \text{sy_gaiamag}$ y $d = \text{sy_dist}$, que ya nos viene dado en parsecs.

También pueden ensayarse otras distribuciones que involucren otros parámetros de los disponibles en busca de posibles correlaciones significativas. En el anexo se ofrece otro ejemplo que contempla la edad estelar.

En el próximo artículo entraremos en un tema mucho más interesante: las relaciones masa/radio de los exoplanetas. Ahí intentaremos discriminar entre planetas rocosos, tipo Tierra, y gaseosos como los gigantes tipo Júpiter y mayores.

ANEXO 2

1- Python para distribuciones de parámetros, relacionados dos a dos, en una parrilla de ejes múltiples.

a) Primero es necesario exportar un fichero .csv desde Excel con el contenido que se quiere representar. En este caso, la mayoría de los campos del fichero, (sin los correspondientes parámetros de errores, etc., para simplificar).

b) Otro gráfico con algunos parámetros diferentes a los de la fig. 2.1:

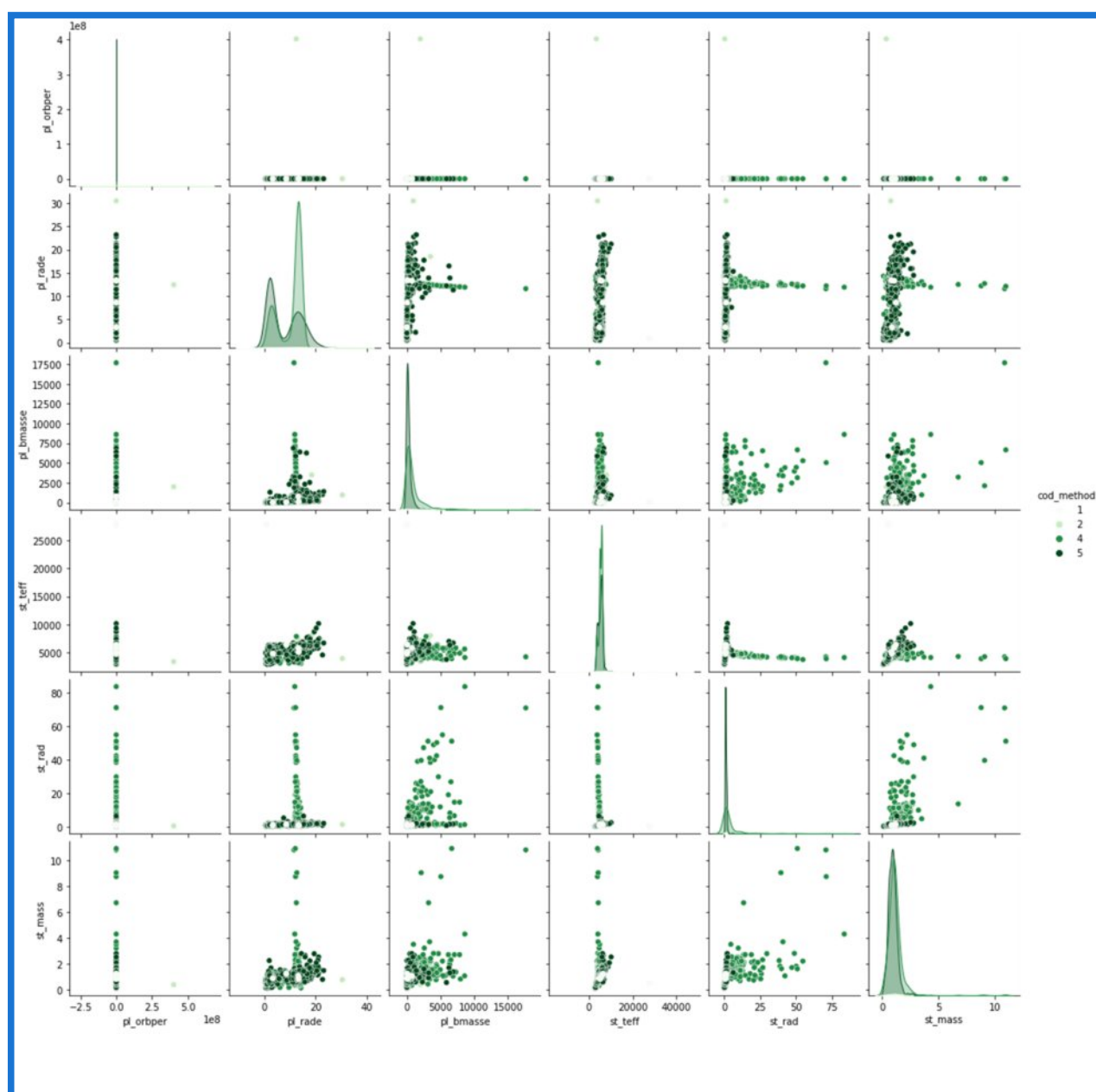


Fig. 2.1b – Correlaciones múltiples por pares de variables

c) Código Python:

librerías

```
import numpy as np
import pandas as pd
import seaborn as sns
```

lectura del fichero .csv exportado (exoplanets-2-correlaciones.csv) y expresión de variables

```
df = pd.read_csv('exoplanets-2-correlaciones.csv', sep=';',
names=['pl_name', 'hostname', 'discoverymethod', 'cod_method', 'pl_orbper',
'pl_orbsmax', 'pl_rade', 'pl_radj', 'pl_bmasse', 'pl_bmassj',
st_spectype', 'st_teff', st_rad', st_mass', 'st_met', 'st_metratio',
'st_logg', 'sy_dist', 'sy_gaiamag'],
skiprows=1) # salta el 1er. registro (cabecera)
df.dropna(inplace=True) # elimina registros con valores nulos
```

gráfico

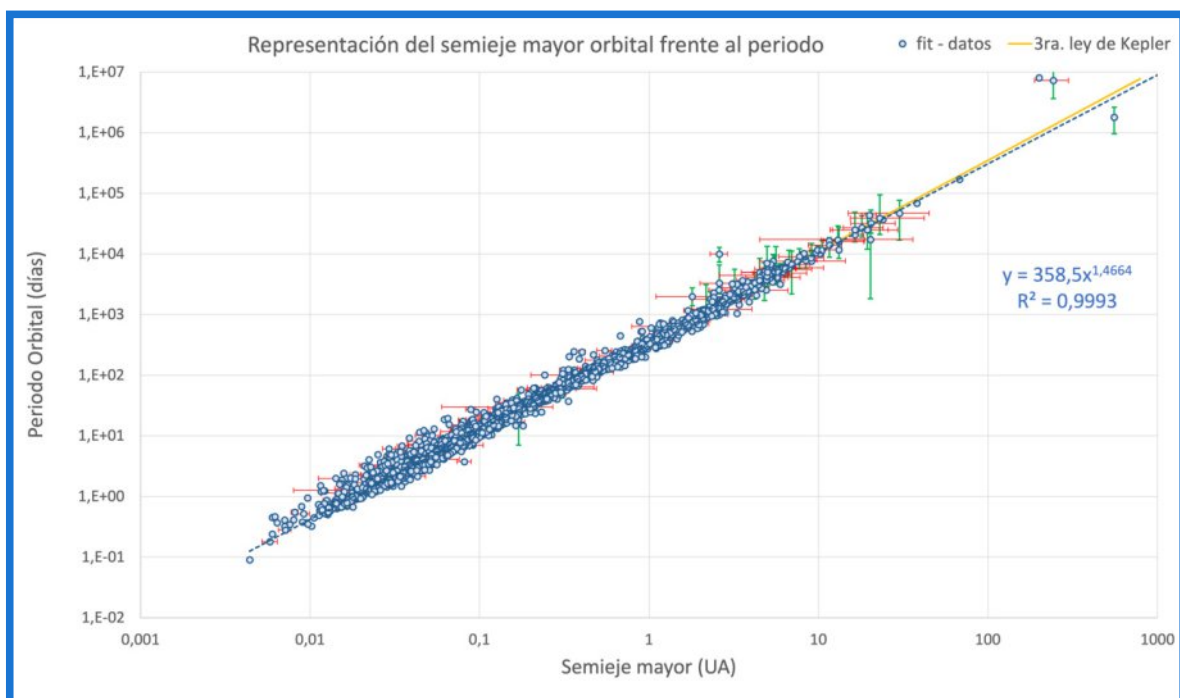
```
sns.pairplot(df, vars=['pl_orbsmax', 'pl_rade', 'pl_bmasse', 'st_logg', 'sy_dist', 'sy_gaiamag'],
hue='cod_method', palette='Greens') # para fig. 2.1 en el artículo
sns.pairplot(df, vars=['pl_orbper', 'pl_rade', 'pl_bmasse', 'st_teff', 'st_rad', 'st_mass'],
hue='cod_method', palette='Greens') # para fig. 2.1b en el anexo
```

2- La tercera ley de Kepler

Si se utilizan las columnas que contienen los errores estimados en las medidas de *pl_orbper* y *pl_orbsmax* y se introducen en la gráfica de la fig. 2.4, se tendrá una imagen más significativa de la calidad de estos datos:

La línea en amarillo superpuesta representa la función $y = C \cdot x^{1.5}$ esto es, la tercera ley de Kepler, (con $C \approx 350$), totalmente compatible con los datos tratados en el margen de confianza de sus incertidumbres empíricas.

Fig. 2.4b – La misma representación que la de la fig. 2.4 pero incorporando la información relativa a los errores recogidos en las columnas correspondientes del fichero de exoplanetas.



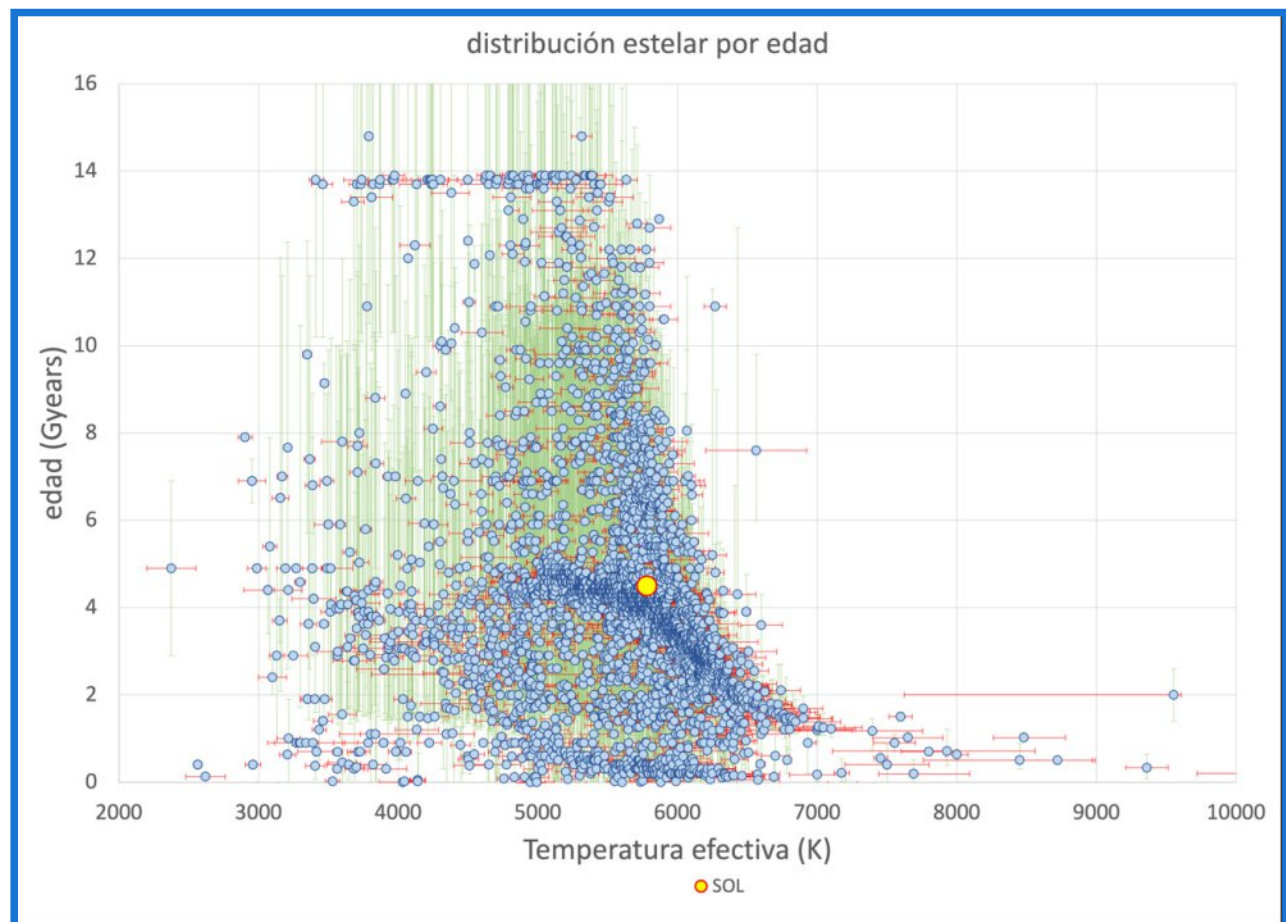
3. Una posible relación entre la edad estelar y la temperatura efectiva de las estrellas.

El objetivo de este ejercicio es llamar la atención sobre la precaución necesaria a la hora de tratar datos de manera masiva y no discriminada. Como se puede apreciar con el parámetro 'st_age' hay un número alto de estrellas con edades elevadas, (próximas a 14 Gyears, e incluso por encima). Este dato es sospechoso en tanto que significaría una edad similar a la del Universo o superior.

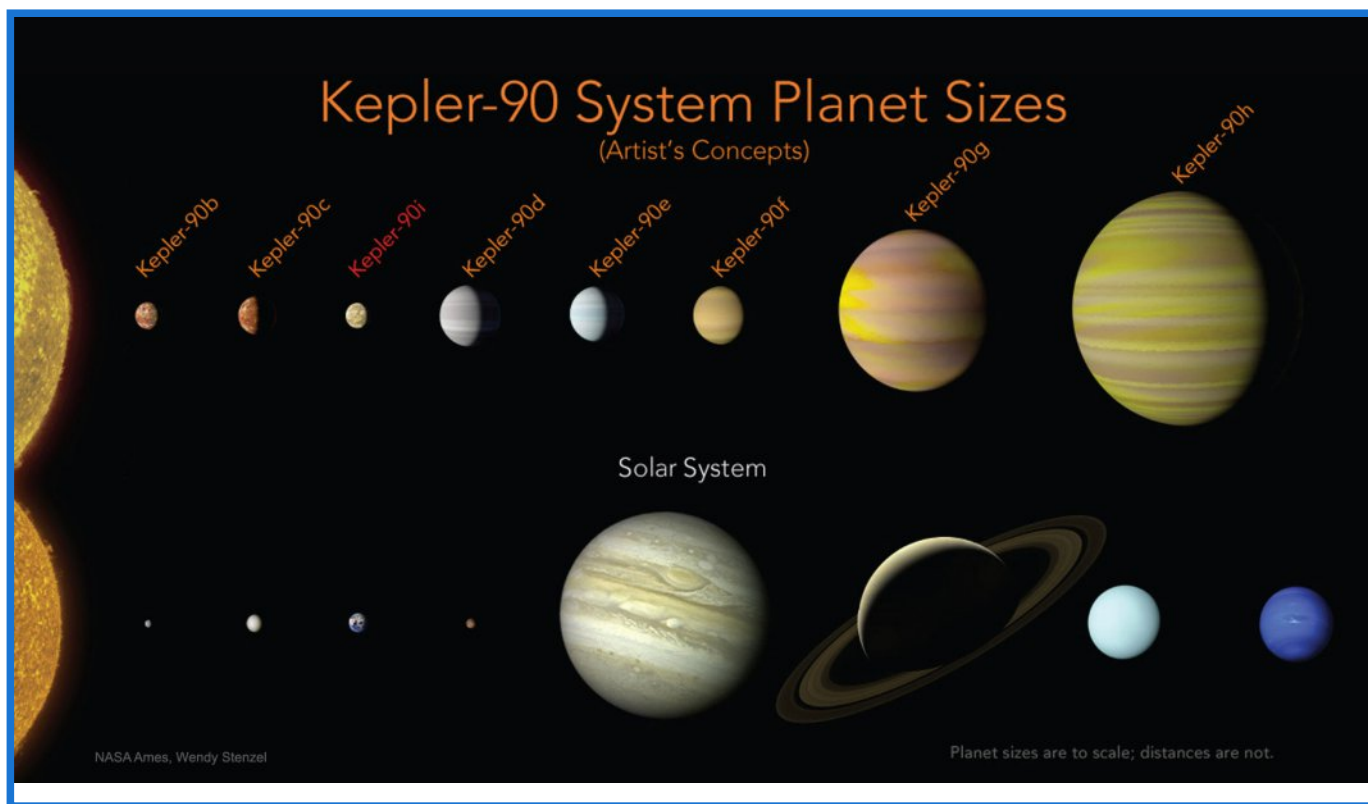
Si en este caso echamos mano de los márgenes de error indicados para estas estimaciones, vemos que la alta incertidumbre que las acompañan explica de alguna manera las anomalías detectadas. Es conveniente, si se quiere hacer algún tipo de estudio con datos de esta naturaleza, restringir los registros útiles a aquellos que presenten unos valores de error por debajo del 15 o 20% del valor absoluto del parámetro tratado.

En el ejemplo aquí presentado parece intuirse un comportamiento con algún significado para edades inferiores a los 6 Gyears, pero es preciso hacer antes una selección con el criterio indicado.

Fig. 2.7 – Representación de la edad estelar en función de la temperatura efectiva.



3ª PARTE



En esta tercera y última parte del trabajo con datos reales a partir del fichero “[Planetary Systems Composite Data](#)”^[1] entraremos en un mayor nivel de detalle con ejercicios que pretenden acercarnos a un mejor conocimiento de lo que una base de datos de estas características, estadísticamente ya muy significativa, puede ofrecernos sobre las relaciones entre la masa, la densidad y el radio.

La relación masa – radio

Una de las propiedades más importantes a considerar en el estudio de los exoplanetas es la relación entre la masa y el radio. No debe olvidarse que estos dos parámetros

están relacionados entre sí y definen su densidad:

$$d = \frac{3}{4\pi} \frac{M}{R^3}$$

^[1] "This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program."

La densidad de un planeta es un indicador de su carácter ‘rocoso’, tipo terrestre, o ‘gaseoso’, tipo Júpiter. Véase, como referencia, las densidades de la Tierra y Júpiter⁽²⁾:

$$d_T = 5,515 \text{ g/cm}^3$$

$$d_J = 1,336 \text{ g/cm}^3$$

Como se puede ver, parece haber una separación clara en el valor de las densidades que reflejaría la estructura y composición material de los planetas. Pero ¿se puede mantener esta caracterización con una base estadística más amplia como la que ofrecen los 5.030 exoplanetas confirmados?⁽³⁾ ¿Hay un umbral nítido que separe los tipos de planetas en función de sus densidades? ¿Puede hablarse solo de planetas rocosos y planetas gaseosos?

Sabemos que una condición importante para que un planeta sea habitable es que éste

retenga una capa atmosférica gaseosa, pero al mismo tiempo parece fundamental que también tenga una base rocosa con agua líquida en superficie. Al margen de lo que esto signifique en cuanto a las condiciones de habitabilidad, caracterizar esta tipología es fundamental también para comprender mejor los mecanismos de formación y evolución de los sistemas planetarios. Estas condiciones por ahora son difícilmente observables con las técnicas de observación y detección disponibles. Tal vez el [telescopio espacial James Webb](#), o la futura [misión Ariel \(ESA\)](#), puedan aportar información más determinante al respecto en los próximos años.

Con los datos disponibles actualmente intentaremos responder a alguna de estas preguntas analizando la distribución de los exoplanetas en base a la relación masa/radio. Empezaremos por representar esta distri-

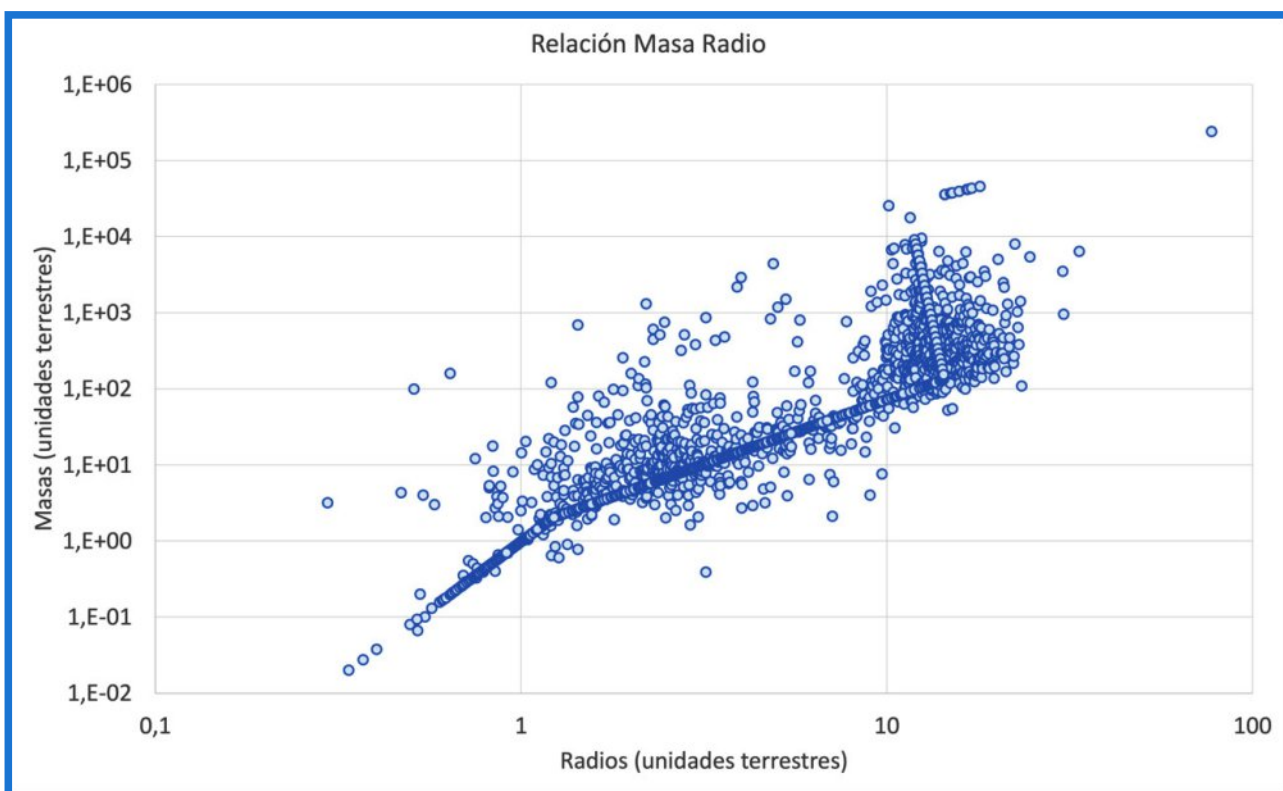


Fig. 3.1 – distribución de exoplanetas en función de sus radios y masas. Ejes en escala logarítmica para hacer más comprensible el gráfico⁽⁴⁾.

⁽²⁾ Otros planetas del sistema solar: rocosos > Mercurio: $d_m=5,43 \text{ g/cm}^3$, Venus: $d_v=5,24 \text{ g/cm}^3$, Marte: $d_M=3,933 \text{ g/cm}^3$; gaseosos > Saturno: $d_s=0,690 \text{ g/cm}^3$, Urano: $d_U=1,274 \text{ g/cm}^3$, Neptuno: $d_N=1,64 \text{ g/cm}^3$

⁽³⁾ A fecha de 9 de mayo de 2022.

⁽⁴⁾ En el anexo puede verse otra representación de este mismo gráfico con el eje de ‘radios’ en escala de valores absolutos, no logarítmico, también en unidades de radios terrestres.

bución haciendo uso de los valores contenidos en las columnas '*pl_rade*' y '*pl_bmasse*', esto es radios y masas de los exoplanetas en unidades terrestres, (fig. 3.1).

Lo primero que llama la atención en esta representación es una línea en *zig-zag* ascendente con tres segmentos nítidos y que aglutinan el 77% de los exoplanetas representados. Este porcentaje se detallará en breve.

Cuando uno se encuentra con una *regularidad* de esta naturaleza, si no está sobre aviso, se corre el riesgo de sufrir una taquicardia, porque algo así puede indicar la posibilidad de haberse topado inesperadamente con algo con un posible significado físico más o menos inmediato. La alternativa es poner en duda la calidad o naturaleza de los datos, especialmente por la ausencia de dispersión de estos en torno a esa línea bien marcada, como correspondería a una serie normal de datos empíricos.

Pero lo que ocurre aquí es fácil de explicar. Mientras que el método de detección por tránsito puede medir con cierta facilidad el radio planetario no ocurre lo mismo con la masa. Por el contrario, el método de *velocidad radial* puede determinar las masas, pero no es capaz de medir radios. De manera que en nuestro fichero "[Planetary Systems Composite Data](#)" tenemos tres tipos de registros:

A. Exoplanetas de los que se conoce empíricamente el radio pero no la masa.

B. Exoplanetas de los que se tiene una medida de la masa pero no del radio.

C. Exoplanetas en los que ambos parámetros tienen una estimación empírica porque han sido observados con dos o más métodos.

Lo primero que vamos a hacer es separar estas tres clases en nuestro fichero. Para ello necesitamos echar mano de la columna '*pl_rade_reflink*' que tiene la información acerca de cómo o quién ha hecho la medida/estimación del radio. Los exoplanetas para los que no se tiene esta medida arrastran en este campo la cadena '*Calculated Value*'.

Paralelamente, los registros que corresponden a exoplanetas para los que no se tiene estimación de su masa, llevan en la columna '*pl_bmassprov*' el valor '*M-R relationship*'.

1. Para facilitar el tratamiento de estas tres clases se recomienda añadir sendas columnas a *pl_rade_reflink* y *pl_bmassprov* con algún indicador sencillo que sustituya las cadenas de caracteres de las anteriores, (por ejemplo, 1 = valores calculados, 0 = valores empíricos). De esta manera tendremos las tres clases identificadas fácilmente:

$$\begin{aligned}(\text{radio}, \text{masa}) &\equiv (0, 1) \rightarrow \text{clase A} \\(\text{radio}, \text{masa}) &\equiv (1, 0) \rightarrow \text{clase B} \\(\text{radio}, \text{masa}) &\equiv (0, 0) \rightarrow \text{clase C}\end{aligned}$$

El caso (1,1) no se da salvo para un puñado escaso de planetas, (ver nota al pie [5] de la página siguiente). Hecho esto, podemos representar de nuevo la fig. 3.1 con las tres series diferenciadas (fig 3.2).

Como ya tenemos separadas las tres clases, podemos contabilizar fácilmente el número de sus respectivos integrantes:

Clase A – 2.785 exoplanetas

Clase B – 1.097 exoplanetas

Clase C – 1.134 exoplanetas⁽⁵⁾

Esto es, de sólo el 22,61 % de los exoplanetas se tienen radios y masas medidos empíricamente, mientras que el 77,39 % tienen alguno de estos dos parámetros calculados por falta de una estimación directa a partir del método empleado en su detección.

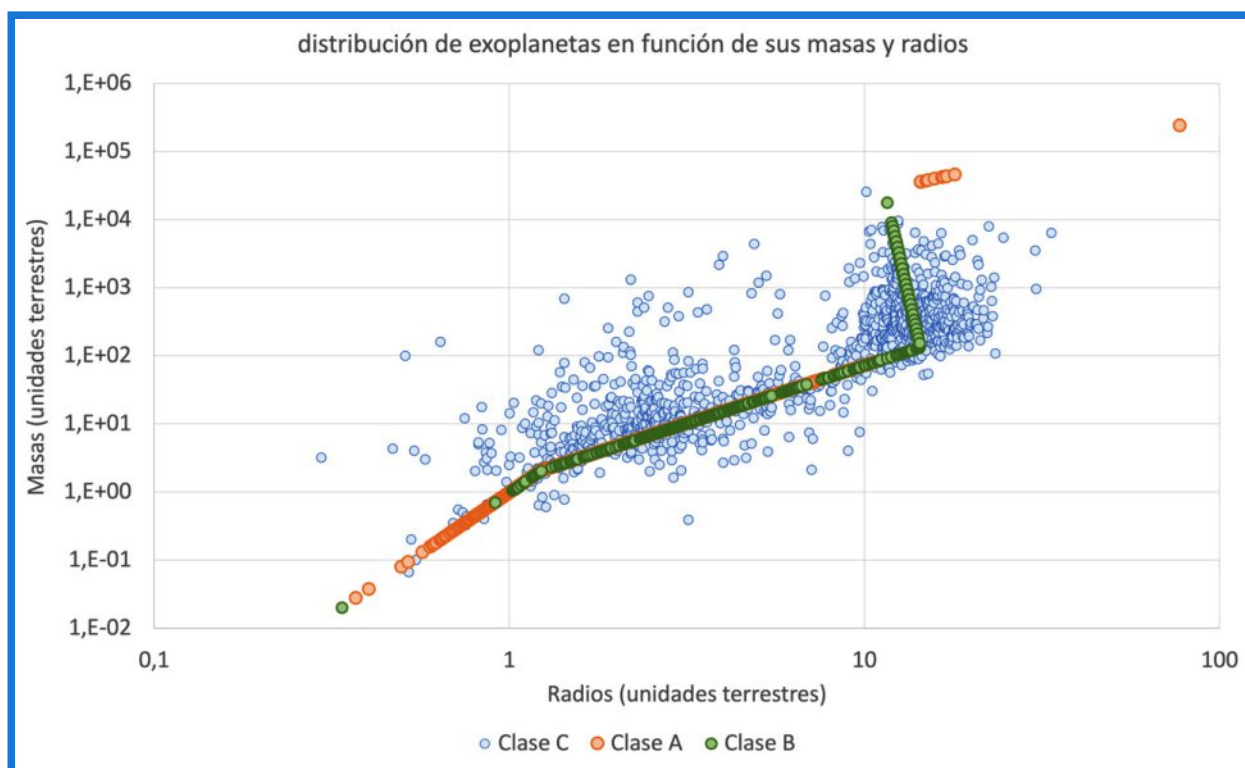


Fig. 3.2 - distribución de exoplanetas en función de sus radios y masas. Ejes en escala logarítmica. La abundancia de la población de la clase A queda parcialmente enmascarada en el gráfico por encontrarse superpuesta sobre ella la población de la clase B.

La pregunta ahora es ¿cómo se han calculado estos valores desconocidos?

Entraremos en ello en breve. Antes será interesante hacer una representación de la distribución de densidades.

Distribución de densidades.

Con las densidades recogidas en el fichero que estamos tratando pasa lo mismo. Hay densidades estimadas empíricamente y otras calculadas. La información en cada caso está presente en la columna 'pl_dens_reflink'. Como para algunos exoplanetas el valor correspondiente a la densidad falta⁽⁵⁾ y en otros está directamente calculado a partir de los radios y las masas, optaremos por añadir a nuestra tabla una nueva columna en la que insertaremos siempre el valor calculado para la densidad, que será el que utilizemos en

adelante para los gráficos correspondientes. Esto no desvirtúa el dato en sí mismo⁽⁶⁾ y puede servir de ejercicio en clase para entender las relaciones masa, radio, densidad y la transformación de unidades relativas en unidades del sistema CGS.

Dado que tanto las masas como los radios están en unidades terrestres, para pasar las densidades (también relativas al planeta Tierra) a unidades de g/cm^3 habrá que multiplicar por $5,515 \text{ g/cm}^3$, que es la densidad de la Tierra en unidades CGS. Por tanto este nuevo campo calculado será:

$$[1] \quad dc_p = \frac{M_p}{R_p^3} \cdot 5,515 \text{ g/cm}^3$$

La distribución de densidades se presenta en la fig. 3.3 a continuación.

⁽⁵⁾ Hay 14 exoplanetas que no tienen informado el radio (de los cuales 5 tampoco informan sobre la masa), ni tampoco ofrecen referencias. Estos 14 registros (todos de la clase C) los retiraremos de los análisis que hagamos desde aquí. Nuestro fichero tendrá finalmente 5016 exoplanetas.

⁽⁶⁾ Ver en el Anexo.

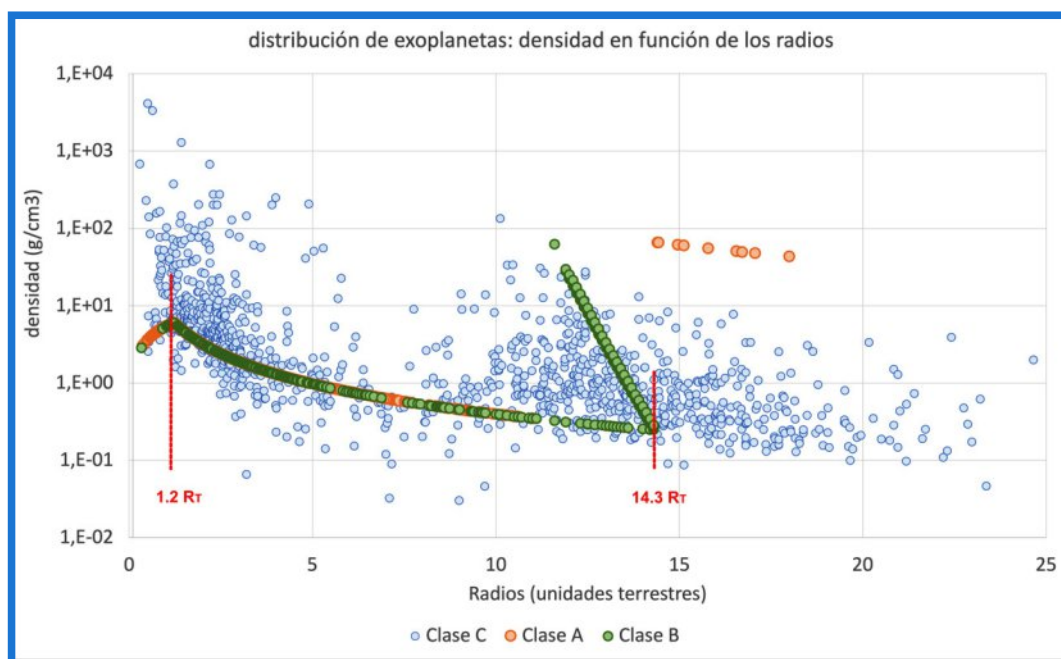


Fig. 3.3: Densidad frente a radio. Eje de ordenadas en escala logarítmica. Se aprecia igualmente el efecto artificioso de los planetas que aportan radio o masa calculado.

También aquí se aprecian los tres segmentos ya reseñados en las figuras 3.1 y 3.2, siendo mucho más llamativos los puntos de inflexión (o ruptura) que aparecen en torno a $1.2 R_T$ y $14.3 R_T$.

Ya es el momento de hablar sobre la forma de asignar valores calculados a la masa cuando sólo se conoce el radio y viceversa que, como queda claro, debería explicar la distribución de ese 77,39 % de exoplanetas en cualquiera de las distribuciones presentadas hasta aquí.

El algoritmo de 'Chen & Kipping'⁽⁷⁾.

Se trata de un procedimiento desarrollado por Chen y Kipping a finales de 2016 en base a la aplicación de una estadística bayesiana⁽⁸⁾ sobre los exoplanetas de la clase C, es decir, aquellos que sí tienen medidas para los radios y las masas. A partir de los parámetros derivados de este análisis se proyectan los valores desconocidos, masas para la clase A y radios para la clase B.

A pesar de que el algoritmo en sí, que ahora veremos con más detalle, extrapola los resultados que nos han llamado la atención y que podríamos considerar como poco naturales, es el procedimiento que aún se sigue utilizando para resolver el problema planteado. En 2016 el número de exoplanetas que formaban lo que hemos llamado clase C era sólo de 316, y ésta fue la base estadística que da lugar al algoritmo. Posteriormente, con una base extendida a una clase C más amplia, este tipo de método para estimar un mejor algoritmo se mantiene más o menos inalterado, salvo pequeñas correcciones en las posiciones decimales del orden de las centésimas en las leyes de potencia que lo definen.

El algoritmo, que puede verse en la [documentación del fichero](#) que estamos tratando, asume que la relación entre la masa y el radio de los planetas debe ser del tipo de la ley de potencia siguiente:

$$[2] \quad R_p = Cte \, M_p^S \Leftrightarrow M_p = Cte' \, R_p^{1/S}$$

⁽⁷⁾ "Probabilistic Forecasting of the Masses and Radii of Other Worlds" - Jingjing Chen, David M. Kipping – nov 2016

⁽⁸⁾ A las estadísticas bayesianas se las suele considerar como 'estadísticas de probabilidad esperada'. Un ejemplo de problemas que se enfocan bajo estos modelos puede ser: si cuando tengo gripe tengo una probabilidad xx de tener fiebre, cuál será la probabilidad de que, si tengo fiebre, ésta se deba a que tengo gripe.

Si tomamos logaritmos en las expresiones anteriores, tendremos:

$$R = C + M S \Leftrightarrow M = (R - C)/S$$

siendo $M = \log_{10} M_p$, $R = \log_{10} R_p$ y (C, S) los parámetros a determinar por la estadística.

Una vez ajustados estos, los valores que definen el algoritmo son:

$$[3] \quad \{C, S\} = \begin{cases} \{0.00346, 0.2790\} & \text{if } R_p < 1.23 R_T \\ \{-0.0925, 0.589\} & \text{if } 1.23 \leq R_p < 11.1 \\ \{-2.85, 0.881\} & \text{if } R_p \geq 14.3 R_T \end{cases}$$

Si se toman las masas como referencia, la equivalencia de las expresiones [2] se traduce en:

$$[4] \quad \{C, S\} = \begin{cases} \{0.00346, 0.2790\} & \text{if } M_p < 2.04 M_T \\ \{-0.0925, 0.589\} & \text{if } 2.04 \leq M_p < 132 \\ \{1.25, -0.044\} & \text{if } 132 \leq M_p < 26600 \\ \{-2.85, 0.881\} & \text{if } M_p \geq 26600 M_T \end{cases}$$

A simple vista éste resulta un algoritmo feo, ya que el tramado (visible en las representaciones de las figuras 3.1, 3.2 y 3.3), está presente en el mismo desde su propia base y los ajustes correspondientes a los parámetros $\{C, S\}$ asumen los límites marcados en los condicionantes respectivos como zonas donde es posible encontrar correlaciones razonables, independientes y no aplicables fuera de dichos límites. Ver en el anexo para mayor detalle, (fig. 3.3.1 y 3.3.2).

Al margen de lo que nos parezca, éste es el procedimiento que se sigue aplicando para asignar los valores desconocidos (de radios y masas) a falta de otro mejor. Pero lo importante ahora es centrarnos en su interpretación y significado.

A partir de los rangos de aplicación del algoritmo [3], que define tres umbrales para los radios: (1.23, 11.1 y 14.3 R_T respectivamente), el espectro de distribución de los planetas se fragmenta en cuatro zonas (1, 2, 3 y 4) a las que los propios Chen y Kipping asignan una posible interpretación. Así, los exoplanetas con $R_p < 1.23 R_T$ serían *planetas rocosos*, los comprendidos en el intervalo $1.23 \leq R_p < 11.1 R_T$ los consideran gaseosos '*tipo Neptuno*' y para aquellos cuyo radio está por encima de $14.3 R_T$ reservan la posibilidad de mundos estelares, o sea proto-estrellas en las que es posible que se haya iniciado la combustión de hidrógeno, (pero sólo para los de mayor densidad en esta franja).

La zona 3, sólo contemplada en [4] y que alojaría exoplanetas con masas en el intervalo $132 \leq M_p < 26600 M_T$, es considerada propia de los planetas *gaseosos 'tipo Júpiter'*, (ver fig. 3.5). Como puede verse, en esta región el algoritmo tiene un comportamiento singular ya que parece haber un retroceso en el tamaño del radio a medida que aumentan las masas. Esto, físicamente, puede interpretarse fácilmente, puesto que la acumulación de masa (gas), produce un aumento del potencial gravitatorio del planeta, que se contrae aumentando drásticamente su densidad. A esta región se le llama de '*radio degenerado*', debido a la multiplicidad de valores que puede tomar la masa para un mismo valor del radio, (véase que la serie de exoplanetas proyectados aquí dibujan una línea que se acerca a la vertical en el gráfico).

En este último gráfico, donde en el eje de abscisas se representan masas, vemos que en la zona 3 (planetas tipo Júpiter y mayores), la masa sigue creciendo como era de esperar a pesar de que los radios se contraigan como se ve en la fig. 3.4.

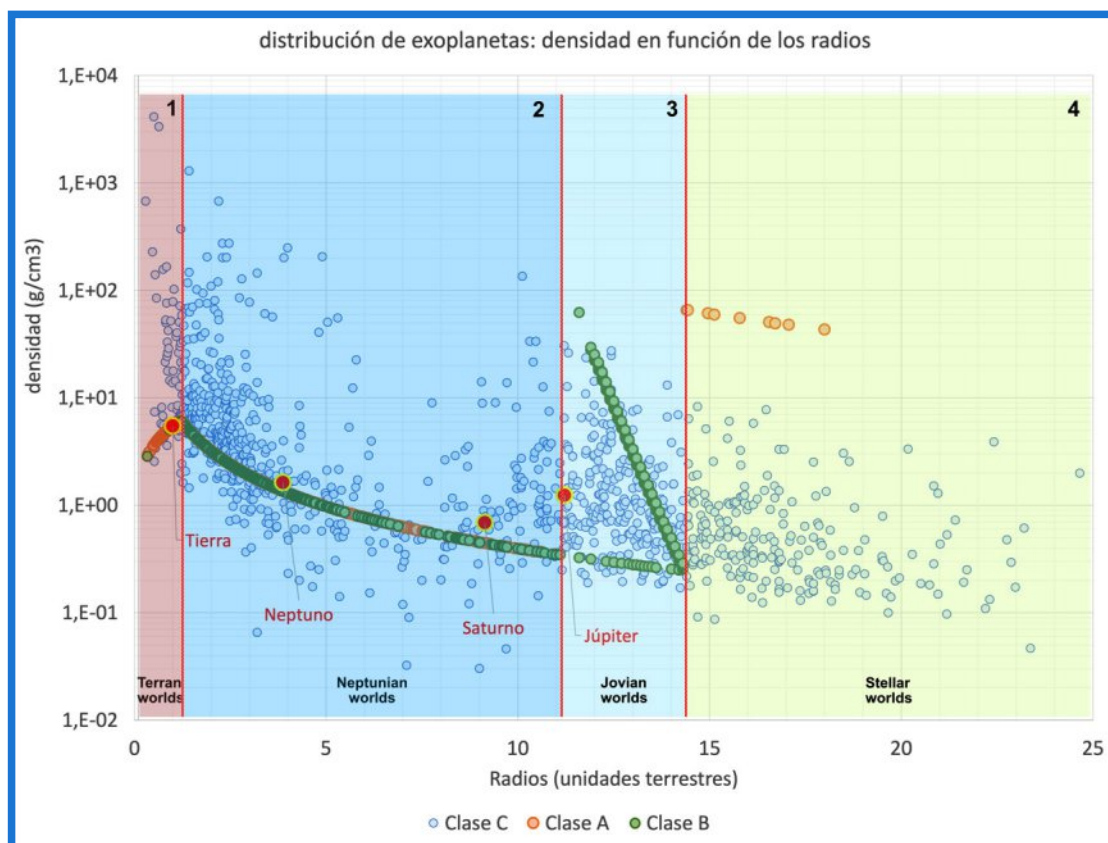


Fig. 3.4 Es la misma representación de la fig. 3.3 pero se han añadido algunos elementos para su mejor interpretación. También se han incorporado las posiciones de los planetas Tierra, Júpiter, Saturno y Neptuno, como referencias interesantes a considerar.

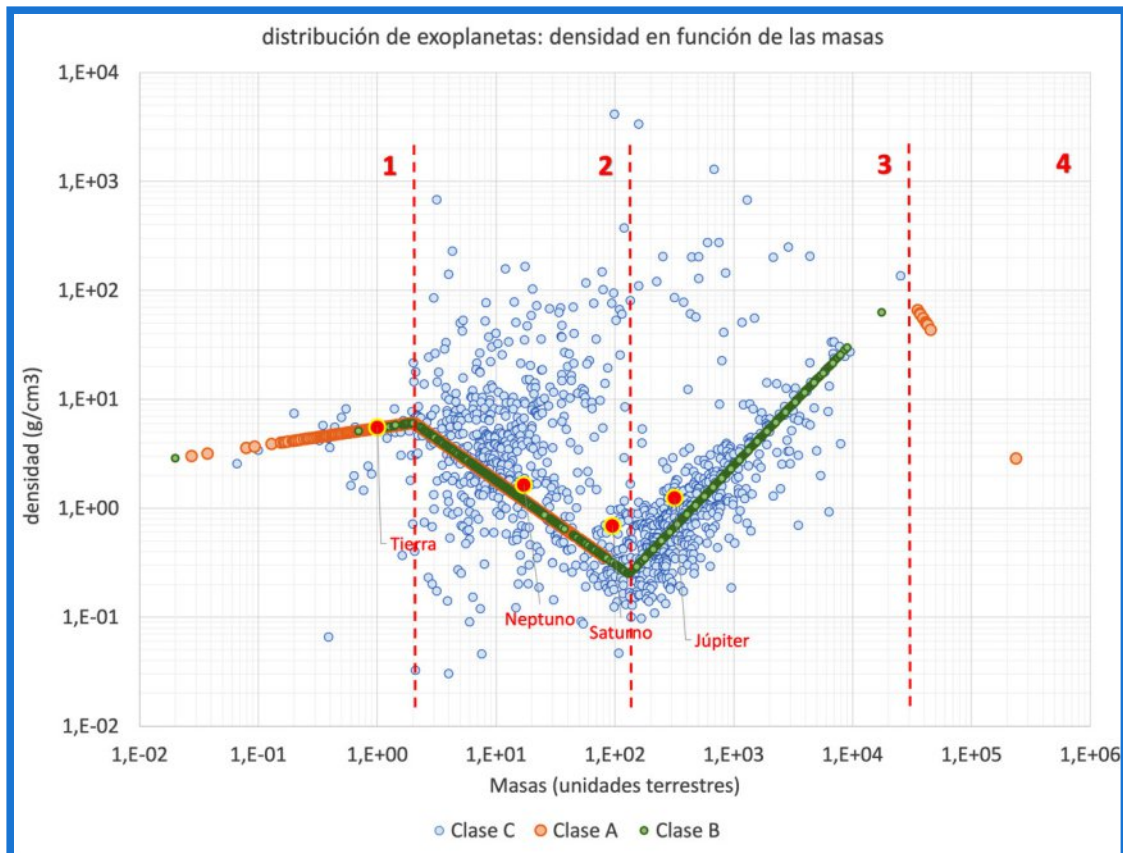


Fig. 3.5 – Distribución de densidades en función de la masa.

Lo más significativo, no obstante, es el punto de inflexión que separa los planetas rocosos de los gaseosos tipo Neptuno, en valores muy próximos a los que tiene el planeta Tierra ($1.23 R_T$, $2.04 M_T$). Significa esto que, bajo las premisas derivadas del *algoritmo de Chen & Kipping*, la Tierra sería un *gran planeta* rocoso casi en el límite de la región de los gaseosos tipo Neptuno (ver fig. 3.4). Incluso los propios Chen y Kipping vienen a decir en las conclusiones de su artículo "*nuevamente (se) argumenta a favor de la escasez de Súper-Tierras. Puede ser, entonces, que La Tierra es la Súper-Tierra que hemos estado buscando todo el tiempo*"⁽⁹⁾.

Más adelante discutiremos todo esto un poco más en profundidad. Por ahora puede parecer un poco confuso las diferentes representaciones hechas sobre unos mismos datos, alternando ejes y/o escalas logarítmicas con naturales, pero es totalmente intencionado. Primero, porque normalmente un único gráfico no da las diferentes perspectivas que los datos en sí pueden presentar. Por otra parte, hay un objetivo subyacente en estos ejercicios propuestos: el de manejar con cierta soltura las herramientas de análisis empleadas y también, el de ser conscientes de que la interpretación de estos análisis a veces requiere verlos desde diversos puntos de vista.

Modelos mejorados. Alternativas.

A pesar de lo comentado sobre el modelo propuesto por Chen y Kipping, y que éste fuera desarrollado sobre una base estadística todavía escasa (316 planetas confirmados en su momento), este modelo es el que se sigue utilizando todavía hoy como el menos malo (o único posible *hasta nueva orden*) para proyectar la variable desconocida en cada caso, dado que el método basado en la estadística bayesiana mantiene los parámetros sin apenas variación a pesar de

haber aumentado el número de exoplanetas significativamente.

De todas formas hay que resaltar algo que parece evidente en los gráficos que ya hemos desarrollado aquí:

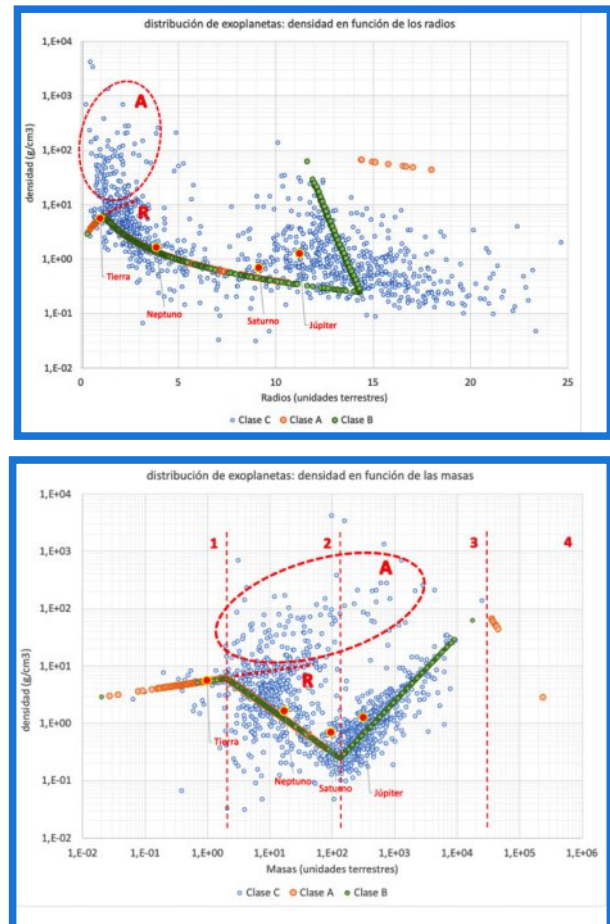


Fig. 4.1 – gráficas 3.3 y 3.5 con dos anotaciones marcadas como R y A

Como se puede contemplar en ambas imágenes, la rama de planetas rocosos parece tener continuidad más allá del drástico límite impuesto por el algoritmo Chen & Kipping, (línea R), llegando hasta planetas con radios próximos a $4 R_T$ (figura superior) y masas del orden de las $40 M_T$ (figura inferior).

Por otra parte, el modelo comentado no dice nada de la nube de exoplanetas de altas y muy altas densidades, marcados en la zona A. Tal vez habría que pensar aquí en planetas con

⁽⁹⁾ "This is also compatible with our determination and again argues for effectively a paucity of Super-Earths. It may be, then, that the Earth is the Super-Earth we have been looking for all along" – Chen & Kipping - "Probabilistic Forecasting of the Masses and Radii of Other Worlds" - nov 2016

una gran atmósfera gaseosa pero con un núcleo rocoso también grande y muy denso. En cualquier caso, se precisan sendas funciones de estado para estas configuraciones, así como esperar a que observaciones más precisas aporten información más determinante.

Mientras estas cuestiones se aclaran, podemos presentar algunos resultados a partir de la aplicación de técnicas de inteligencia artificial (AI) basadas en *machine learning* sobre correlaciones múltiples con los datos existentes a la fecha.

Exoplanetas e Inteligencia Artificial

Ya escapa de los propósitos de esta serie de artículos ensayar en el aula las técnicas de análisis propuestas en este apartado pero, para completar todo lo expuesto aquí, se dan unas breves pinceladas de los resultados obtenidos con estas metodologías recogidos en un artículo de Carlos D. Álvaro y yo mismo^[10].

Lo primero que se aprecia como resultado de este análisis bajo el modelo AI desarrollado es una variedad más rica en el espectro de los exoplanetas que la que se deriva del algoritmo

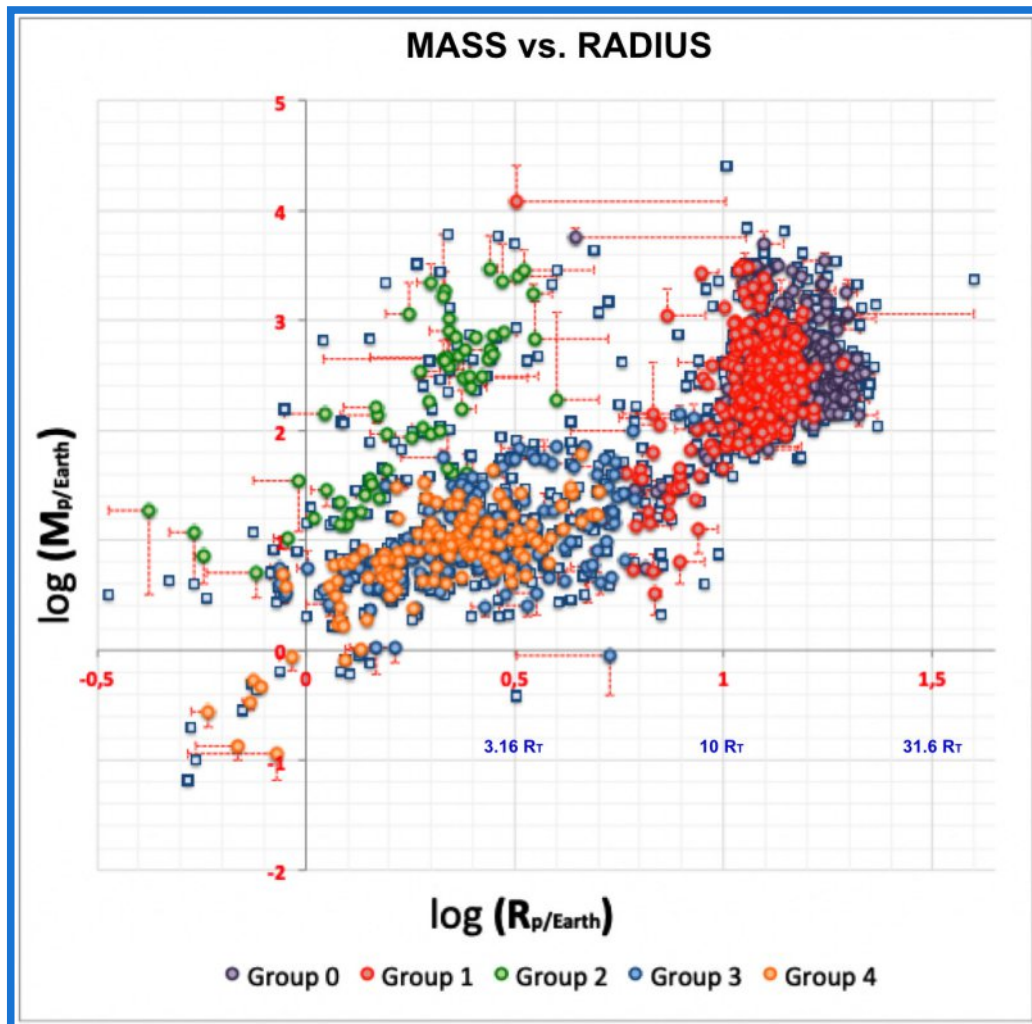


Fig. 4.1 – Representación de masas frente a radios (ejes en escala logarítmica). La figura revela 5 grupos diferenciados después de un *clustering* por *K-Means* sobre el que se aplica posteriormente un modelo de *Random Forest* que incluye la variable 'grupo' del paso anterior.

[10] "Extension of the standard radius-mass relationship of exoplanets using machine learning methods. A population of high-density planets is revealed by machine learning methods in the Planetary Systems Composite Data" – J. Álvaro & Carlos D. Álvaro – Feb 2020 – [Data updated to May 2022].

de Chen & Kipping. Aparecen 5 grupos de los que el grupo 4 puede interpretarse como una población de planetas rocosos, cuyo rango va más allá del límite $1.23 R_T$, $2.04 M_T$, siendo razonable pensar que puede extenderse hasta una frontera próxima a los $4 R_T$ en una región, en el espacio de parámetros (masa, radio), que parecen compartir con planetas de otra naturaleza (posiblemente los llamados ‘planetas de agua’, con un núcleo de hielo comprimido y denso), como zona de transición hacia los planetas gaseosos tipo Neptuno (grupo 3), grupo éste último en el que seguramente tampoco es descartable un núcleo rocoso bajo las grandes envolturas gaseosas.

Los grupos 1 y 0 corresponden evidentemente a los llamados ‘gaseosos tipo Júpiter’ y la separación en dos ramas puede estar diferenciando entre los que acceden a la región de radio degenerado por la izquierda y por la derecha respectivamente.

Pero lo más interesante, a nuestro entender, es la revelación de una región de planetas de pequeño radio y muy densos (grupo 2), sobre los que hay escasa o nula literatura científica y cuya estructura interna es sin duda un reto para los modelos físico-matemáticos que deben explicar tanto su composición material como su génesis.

Conclusión

La conclusión finalmente es que, a pesar de que el número de exoplanetas conocido es ya notorio, se sigue sabiendo poco acerca de la rica variedad vislumbrada sobre las estructuras internas, su composición material y los procesos de formación inherentes. Los métodos actuales tienen un sesgo, ya

comentado, que favorecen la detección de planetas muy próximos a sus estrellas anfitrionas y/o planetas muy masivos, quedando un rango amplio del espectro por cubrir (véase como ejemplo la figura 3.1b en el anexo, donde la franja entre 5 y $10 R_T$ se encuentra poco poblada).

Por otra parte, la falta de información sobre la naturaleza de las atmósferas ayuda poco para responder estas preguntas. Esta falta de conocimiento fundamental justifica que el esfuerzo a medio plazo se esté poniendo en nueva instrumentación más sensible, con espectrógrafos incorporados, que puedan aportar información sobre la composición de los gases de dichas atmósferas. La misión *Ariel*, que será lanzada en 2029, pretende contar antes de su lanzamiento con un inventario de *exoplanetas objetivo*, cuyo estudio ayude a resolver en lo posible parte de estas incógnitas. El proyecto *ExoClock*, que solicita de manera especial la colaboración de observadores amateurs y facilita herramientas para ello, pretende definir en este periodo un conjunto de unos 1.000 exoplanetas que integren ese inventario de *planetas objetivo*.

Con este artículo terminamos la serie dedicada a los exoplanetas. Al margen de que los contenidos pueden considerarse en sí mismos propios de la divulgación en un área muy activa de la Astronomía, se ha pretendido ofrecer un repertorio de ejercicios, que son fáciles de desarrollar en el aula con herramientas sencillas y que no agotan otros posibles. Lo importante aquí es poner de manifiesto que los datos con que trabajan los investigadores están accesibles sin restricciones y que son un excelente material para introducir a los alumnos en el método científico, dado además su marcado carácter didáctico.

ANEXO 3

1. Otra representación de la distribución masa/radio de los exoplanetas.

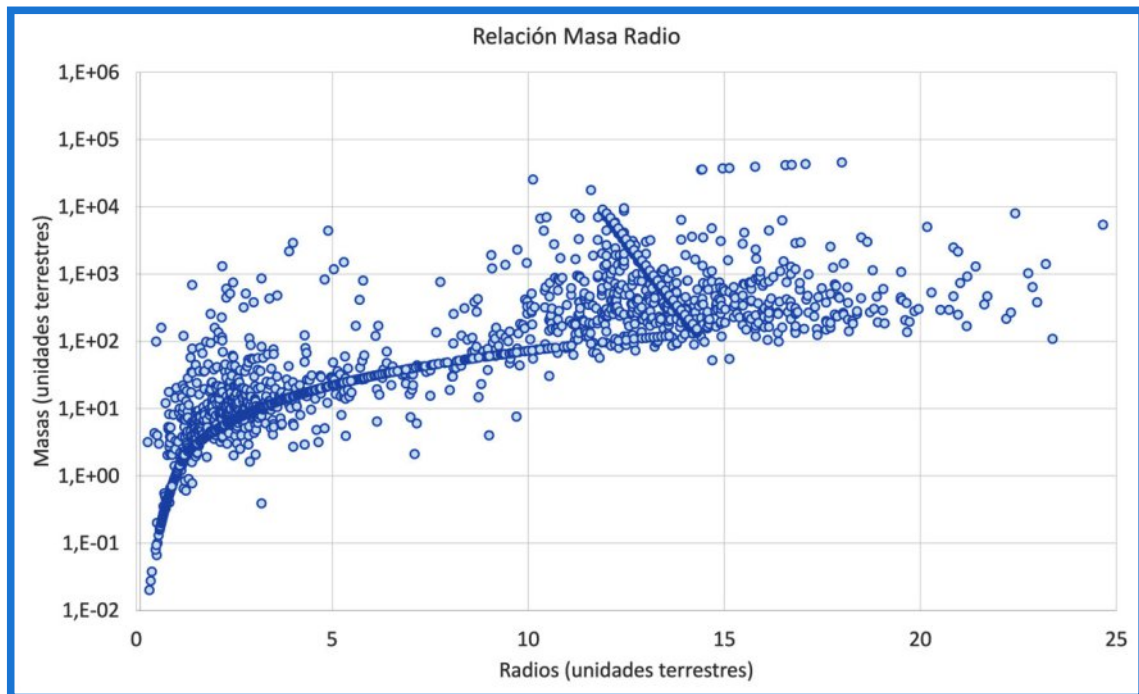


Fig. 3.1b – distribución de exoplanetas en función de sus radios y masas. Eje vertical en escala logarítmica y eje horizontal en unidades naturales. Además de la concentración en una línea sin dispersión, ya apreciada en la fig. 3.1, se hacen más visibles aquí dos poblaciones diferenciadas: a) con radios $< 5R_T$ y b) con radios $10 R_T < R < 20 R_T$

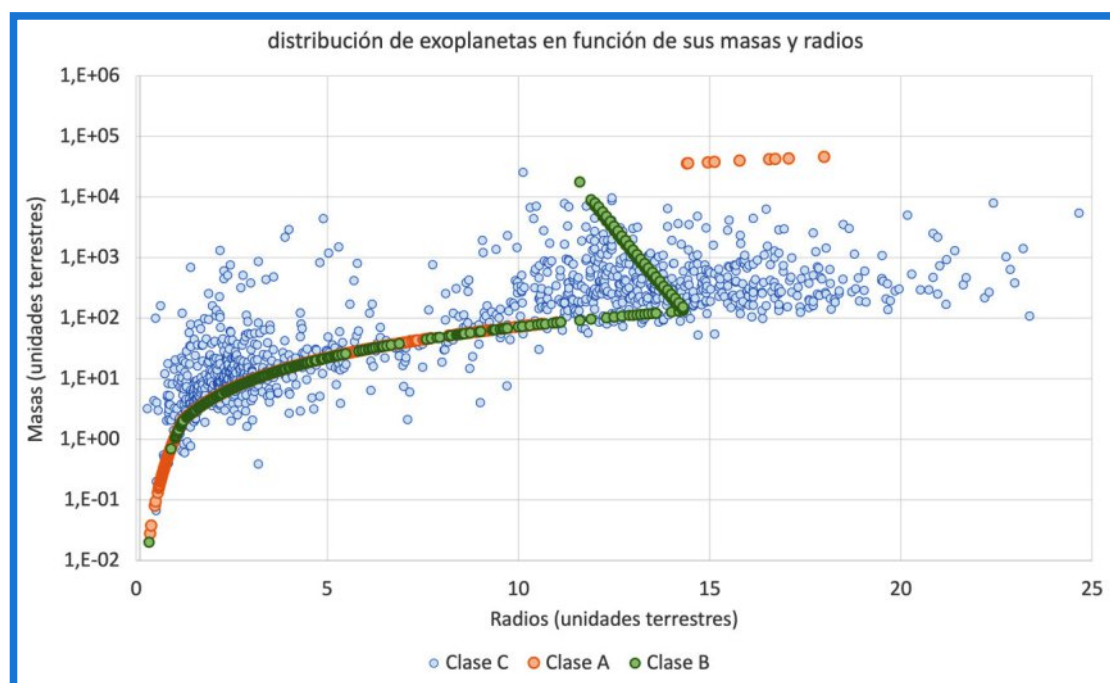


Fig. 3.2b – Distribución de exoplanetas en función de sus radios y masas. Eje vertical en escala logarítmica y eje horizontal en unidades naturales. Se diferencian en colores las respectivas clases A, B y C.

2. Comparación entre las densidades informadas en el fichero de referencia y las densidades calculadas según [1] a partir de los radios y las masas del fichero.

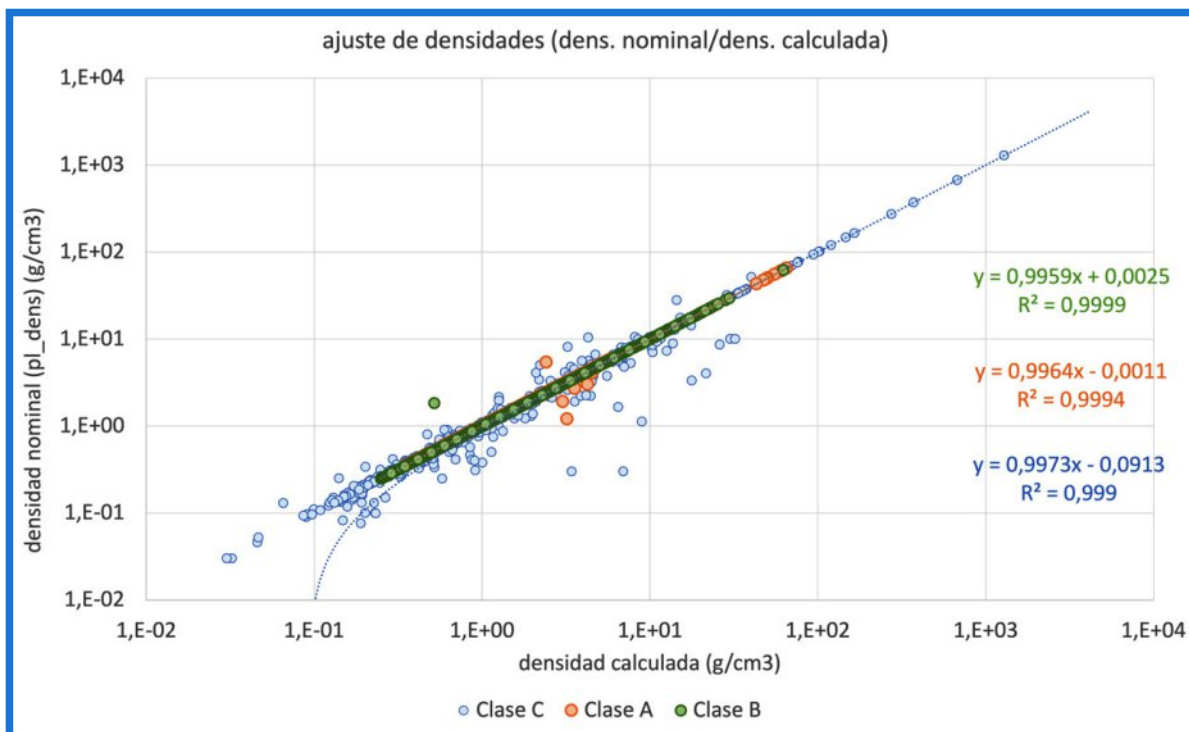


Fig. 3.3b – el ajuste entre los valores calculados para las densidades y los valores nominales presentes en el fichero 'Planetary Systems Composite Data' indica que es muy buena la aproximación de tratar sólo con los valores calculados, evitando los registros que no tienen el dato empírico y también las incertidumbres en las medidas reflejadas en los planetas de la clase C especialmente, que explican la pequeña desviación sobre $y = x$.

3. Representación de las clases A, B y C por separado.

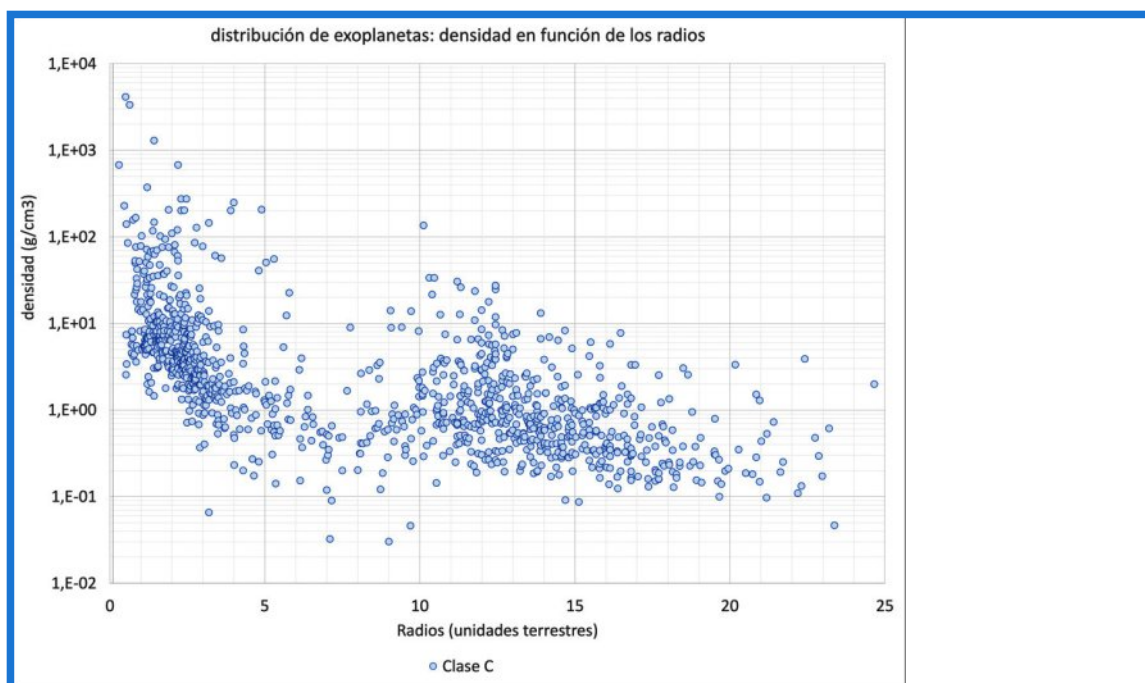


Fig. 3.3.1 – Exoplanetas con medidas en radios y masas. Clase C

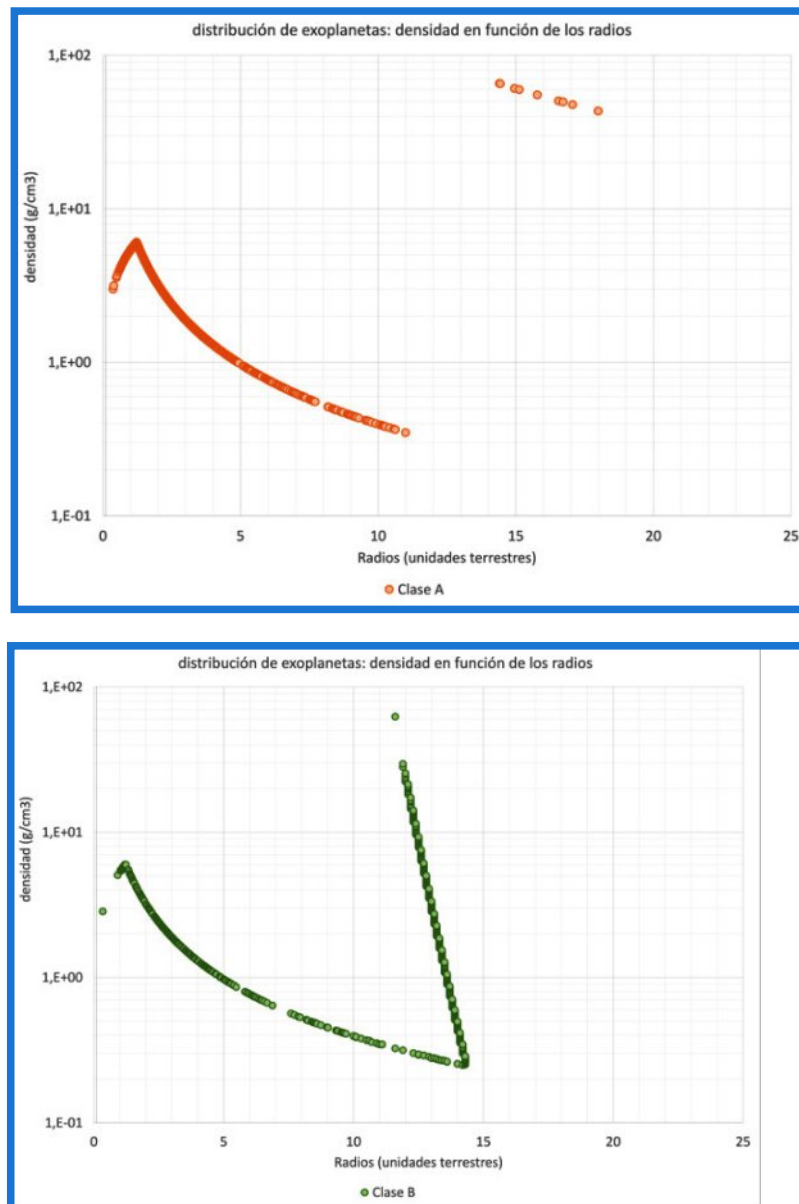


Fig. 3.3.2 – Representación de las clases A y B. En el primer caso se tiene una estimación empírica del radio y la masa se calcula a partir de éste con el algoritmo Chen & Kipping. En la parte derecha de la figura se representa la clase B, (masa conocida y radio calculado).

Estas tres figuras ilustran perfectamente lo que es una distribución natural (fig. 3.3.1) con datos empíricos de las variables radio y masa, frente a la extrapolación calculada según el procedimiento ya comentado para la variable desconocida. Básicamente la clase A corresponde a exoplanetas descubiertos por el método de tránsito, con radio medido, mientras que la clase B la integran planetas detectados por 'velocidad radial'.



ApEA, la Asociación para la Enseñanza de la Astronomía, nació en 1995 para acoger a todas las personas que se dedican a la enseñanza de la Astronomía en centros educativos, planetarios, museos de la ciencia, agrupaciones de aficionados y clubes de estudiantes.

ApEA engloba a todos los interesados en la enseñanza de todos los niveles educativos reglados -desde la enseñanza primaria hasta la universitaria- así como los no reglados. También organiza reuniones de formación para sus socios y publica materiales de interés didáctico, como la presente colección. Más información en www.apea.es

