

¿Cómo se estudian mil millones de estrellas? Astroestadística, el 'big data' del espacio

El satélite Gaia, que durante cinco años estará tomando instantáneas de la Vía Láctea, proporcionará un 'petabyte' de información a los científicos. Gracias a la astroestadística y a la inteligencia artificial, se podrán clasificar las mil millones de estrellas que observe, descubrir objetos celestes desconocidos hasta ahora e incluso entender la evolución de nuestra galaxia. La astronomía ha entrado en la fase del 'big data' y los astrónomos necesitan organizar la información para estudiar el universo.

Cristina Sánchez

15/01/2015 - 12:18h



El satélite Gaia estará durante cinco años obteniendo información sobre la Vía Láctea

En 2013, [un enorme satélite](#), provisto de un plano focal que se ha convertido en la cámara más grande lanzada al espacio, despegó en la Guayana Francesa. La **misión Gaia** comenzaba su periplo de cinco años para trazar un **mapa tridimensional de la Vía Láctea con todo lujo de detalles: va a observar mil millones de**



- P U B L I C I D

LO MEJOR DE #HDR2014



Así enseñan en Stanford a vivir el sueño americano
David G. Ortiz



Tras el rastro de Anonymous
Lucía El Asri



¿Por qué las tripas de mi ordenador son verdes?
Álvaro Hernández

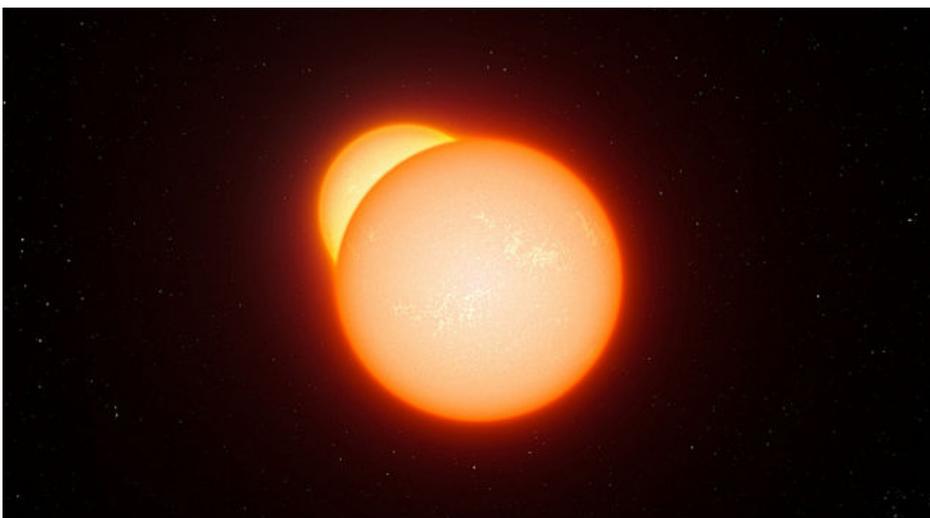
estrellas.

Pero el reto no está solo en observarlas, sino en extraer conocimiento del 'petabyte' de información, un millón de 'gigabytes', que este satélite con nombre de diosa griega va a suministrar. ¿Cómo se gestiona el 'big data' del Universo? **La astroestadística**, la [aplicación de técnicas y modelos estadísticos a la astronomía](#), una [disciplina todavía joven](#), se ha convertido en un **eslabón fundamental para estudiar el 'tsunami' de datos** de asteroides, estrellas, galaxias y cuásares que conviven en nuestro espacio exterior.

HACIA UN 'GOOGLE' DEL ESPACIO

En el año 2000 **se puso en marcha el [Observatorio Virtual \(VO\)](#) con el objetivo de crear una federación de archivos de datos astronómicos**, una única base de datos virtual. La heterogeneidad de archivos dificultaba la labor de los astrónomos, que disponen de un volumen de información sobre el universo cada vez mayor, y el VO nacía para que todos los archivos adoptaran un mismo conjunto de estándares.

Una veintena de países se han unido a esta iniciativa, entre ellos España, que [se sumó en 2004](#). " **La astronomía ha entrado en la fase del 'big data'**. No se puede tener una torre de Babel de archivos astronómicos en la que cada archivo 'hable' su propia lengua y sea imposible establecer una comunicación entre ellos", nos explica **Enrique Solano**, investigador principal del [Observatorio Virtual Español](#) (SVO por sus siglas en inglés).



¿Está acabando la tecnología con la lengua de signos?

[Pilar Chacón](#)



El hotel madrileño que se empeñó en cobrar el wifi al padre de internet

[José Luis Avilés](#)



Tras los pasos de Ana y Mía

[Laura Albor](#)



El póquer online esconde sus ases en mangas extranjeras

[Cristina Sánchez](#)

- P U B L I C I D

L O
+

Leído Comentado



Historia de los inventores españoles que no salen en los libros de texto

[Lucía El Asri](#)

- P U B L I C I D

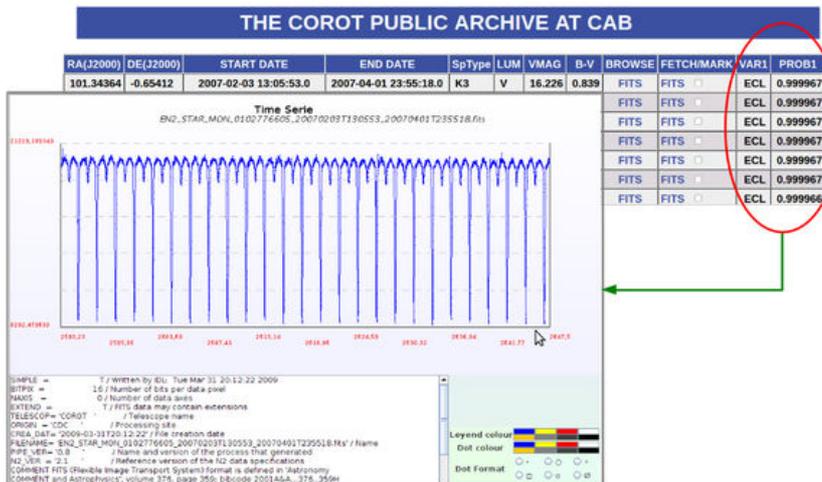
El proyecto del SVO se lleva a cabo desde el [Centro de Astrobiología](#), el instituto que alberga el mayor centro de datos astronómicos de España: **20.000 'gigabytes' de información a disposición de la comunidad científica y del público general.** Ante un volumen de información de tal magnitud, ya no se puede extraer conocimiento por las metodologías habituales. La astroestadística se presenta como la solución a este problema de extracción del conocimiento.

En España, el archivo del [satélite CoRoT](#) recoge un total de 150.000 curvas de luz, es decir, gráficas sobre cómo varía el brillo de una estrella a lo largo del tiempo.

"Hemos desarrollado un sistema de clasificación automática para determinar a qué clase de objeto celeste corresponde cada curva de luz", nos cuenta Solano.

Gracias a este sistema, cuando el astrónomo realice una consulta al archivo de CoRoT, **puede identificar fácilmente los objetos que le interesan** y descargarse solamente lo que necesita. "Antes, si un astrónomo quería identificar en el archivo de CoRoT un tipo particular de objeto variable, por ejemplo, las binarias eclipsantes ([un sistema de estrellas variables que parecen eclipsar una a la otra](#)), **tenía que descargarse esas 150.000 curvas de luz"**, nos cuenta el responsable del Observatorio Virtual en España.

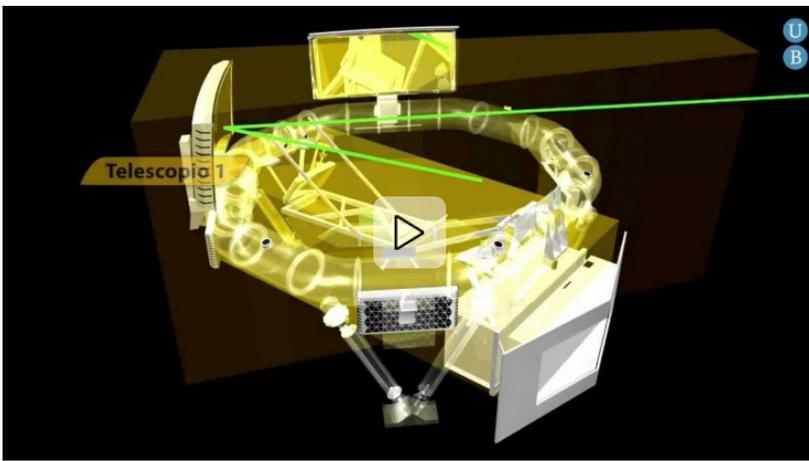
"Ahora, como la información sobre el tipo de objeto se encuentra incorporada al archivo, si un astrónomo busca binarias eclipsantes, podrá descargarse solamente las curvas de luz asociadas a ese tipo de objetos", detalla Solano, que añade que gracias a la astroestadística **los astrónomos pueden abrir nuevas líneas de investigación que antes no eran posibles.**



Salida del archivo de CoRoT en la búsqueda 'binarias eclipsantes'. (Haz clic para ampliar)

Estadísticos, astrofísicos e informáticos trabajan de la mano en esta tarea. **"La prioridad básica es dar acceso a los grandes archivos del mundo, que estén disponibles en el Observatorio Mundial, y esto ya se ha conseguido"**, nos explica Solano. Después, habría que 'exprimir' esos archivos para proporcionar la máxima información posible. Y la aplicación de técnicas de clasificación supervisada al archivo de CoRoT es solamente un ejemplo de lo que la astroestadística puede aportar.

Las ideas y proyectos son numerosos, pero los recursos son limitados. Por el momento, el Observatorio Virtual Español ha centrado sus esfuerzos en el ámbito de la astroestadística en el satélite de la Agencia Espacial Europea Gaia, [una misión que lleva preparándose desde 1998.](#)



¿QUÉ HACER CON MIL MILLONES DE ESTRELLAS?

450 expertos de toda Europa pertenecen al [Consortio de Procesado y Análisis de datos de Gaia](#). **Luis Manuel Sarro**, investigador del departamento de Inteligencia Artificial de la UNED, pertenece a ese consorcio: trabaja en la fase final de la cadena de procesamiento de datos. Su labor se desarrolla entre la astrofísica, la inteligencia artificial y la astroestadística. " **Nunca ha habido ninguna misión que se acercase ni remotamente a la cantidad de datos que va a proporcionar Gaia**; el paso anterior estaba en millones de estrellas", nos explica Sarro.

Este investigador nos cuenta que esta 'aspiradora' espacial observa cada estrella (su posición y sus movimientos) en promedio 70 veces durante los cinco años que durará la misión. Es decir, que **registrará un total de 70 mil millones de observaciones**, cada una de ellas compuesta a su vez de varios conjuntos de medidas. Seis centros de datos repartidos por toda Europa **reciben 50 'gigabytes' diarios de datos de Gaia en inmensos servidores**. Uno de ellos es el [European Space Astronomy Centre](#) de Villanueva de la Cañada, en Madrid.

Superordenadores [del Barcelona Supercomputing Center](#) han contribuido al proyecto realizando distintas simulaciones en el superordenador MareNostrum, que se encargará de extraer y reducir los primeros datos de la misión.



Sarro lleva ocho años trabajando para crear los algoritmos que permitan que la información de Gaia pueda ser clasificada, analizada y transformada en conocimiento sobre nuestro universo. "Hemos necesitado nuevas técnicas no utilizadas antes. **Cada objeto, estrella o cuásar nos ofrece muchísima información, como los usuarios de internet**", nos cuenta.

"Nuestro cerebro se ha especializado evolutivamente en el análisis de datos bidimensionales o tridimensionales (los que nos proporciona nuestra visión espacial), pero **la información astronómica se despliega en muchas más dimensiones, por lo que su análisis requiere de técnicas propias de la inteligencia artificial y de la estadística**", detalla este investigador.

Esas técnicas de minería de datos permiten clasificar esos mil millones de estrellas de forma automática, ya que se ha 'entrenado' a los ordenadores para que sean capaces de **reconocer los objetos celestes y tipificarlos**. "Por ejemplo, son capaces de distinguir entre objetos ultrafríos - a los que no podemos llamar estrellas porque no han llegado a encender el horno nuclear que hay en el centro de una estrella, y que van enfriándose poco a poco - de otras estrellas más calientes", nos detalla Sarro. Conocer datos completos de cada estrella, las propiedades físicas de cada objeto celeste, es uno de los pilares de la misión.

Los algoritmos diseñados para analizar los datos de Gaia también permitirán descubrir objetos celestes nuevos: los superordenadores analizarán por sí mismos los

datos del universo que Gaia les envíe para **investigar tipos de objetos que, hasta el momento, no se conocen.**



El parasol de Gaia durante las pruebas de despegue

"Cuando estemos seguros de que el 'software' funciona bien y está clasificando correctamente los objetos celestes, iremos a analizar la novedad: qué grupos de estrellas o galaxias no encajan en las clases de objetos que conocemos", nos explica Sarro. "Esto va a ocurrir, y es uno de los aspectos más interesantes de la misión.

Gaia nos va a ayudar a descubrir nuevas clases de objetos de nuestra galaxia que no conocíamos. Esto puede suceder por diversas razones. Por ejemplo, porque representen etapas de la vida de una estrella de muy corta duración y, por ello, difíciles de observar", prosigue este investigador.

Además de organizar los datos de ese 'big data' universal, la astroestadística permitirá desvelar otros secretos con los datos de Gaia que no son precisamente baladíes. **"Gracias a los modelos estadísticos, podremos conocer la evolución en el tiempo de la Vía Láctea y su estructura actual** con un nivel de detalle sin precedentes", nos cuenta Sarro.

Esto quiere decir que, además de un 'Google Maps' de la Vía Láctea, la astroestadística permitirá crear una 'timeline' completa de su evolución gracias a los datos de Gaia. **Un paso de los datos al conocimiento, según Sarro.** Por el momento, la misión está en la fase de 'commissioning': los servidores están recibiendo los

primeros datos. Se calcula que en 2021 se publicará el [catálogo final](#). Un impresionante **libro abierto para el estudio de los astrónomos**.

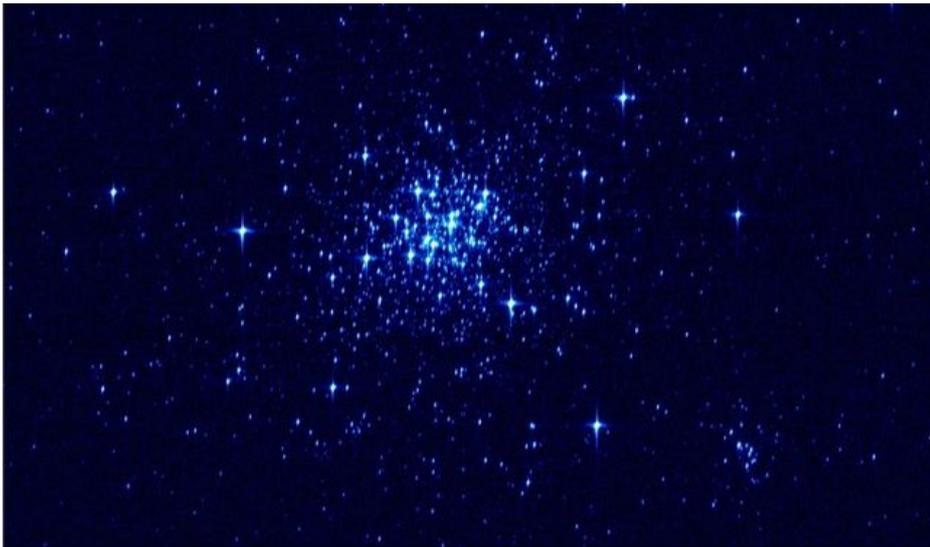


Imagen de calibración que muestra un denso cúmulo de estrellas en la Gran Nube de Magallanes, una galaxia satélite de nuestra Vía Láctea

ESTADÍSTICA Y ASTRONOMÍA, UNA RECIENTE AMISTAD

El término **astroestadística** [nació en la Universidad Estatal de Pennsylvania](#) hace 25 años. Un estadístico y un astrónomo organizaron una serie de conferencias y crearon el [Centro de Astroestadística](#). Allí estudió **Elizabeth Martínez**, física espacial de formación y profesora en el departamento de estadística del [Instituto Tecnológico Autónomo de México](#), una de las pocas profesionales de la joven disciplina en Latinoamérica.

Partiendo de datos de archivos astronómicos públicos, esta investigadora está trabajando en distintos proyectos para analizar, a través de métodos estadísticos, las características de los exoplanetas (que orbitan una estrella diferente al Sol) o estudiar las galaxias. "Hay un problema en astronomía muy antiguo que sigue sin resolverse, **que es la clasificación de galaxias**, y la astroestadística puede ayudar a solucionarlo", asegura.

"La estadística no tiene que verse como una herramienta de análisis, sino como una ciencia necesaria para el estudio de la astronomía", comenta Martínez, que aboga por la convivencia de ambas disciplinas. "Nosotros no empezamos la historia. Algunos padres de la estadística del siglo XVIII también se dedicaban a la

astronomía. La diferencia es que, en el siglo XXI, el **'tsunami' de datos ha obligado a la astronomía a abrazar la estadística como algo imprescindible.** Ya no se puede hacer investigación en astronomía sin utilizar estas técnicas", concluye Sarro.

El **'big data'** [controla nuestras vidas](#), pero también puede servir a fines mucho más trascendentales para la humanidad, como [descubrir el origen y el funcionamiento del universo](#). Todos nos hemos preguntado alguna vez de dónde venimos.

Las imágenes de este reportaje son propiedad, por orden de aparición, de [ESA/ATG medialab](#), [European Southern Observatory](#), Enrique Solano, [ESA-M. Pedoussaut](#) y [ESA/DPAC/Airbus DS](#)



MÁS EN EL DIARIO.ES



¿Puede la Policía pedirte la contraseña del móvil como dijo...



La votante española es de Cuenca (y de derechas)



El teniente Segura es enviado de nuevo a prisión para...



Oscars 2015: 'Birdman' y 'Gran Hotel Budapest' en cabeza...

TAMBIÉN EN LA RED

Nunca es tarde para aprender.



¿Es posible aprender un nuevo idioma si ya eres mayor de... (Babbel)



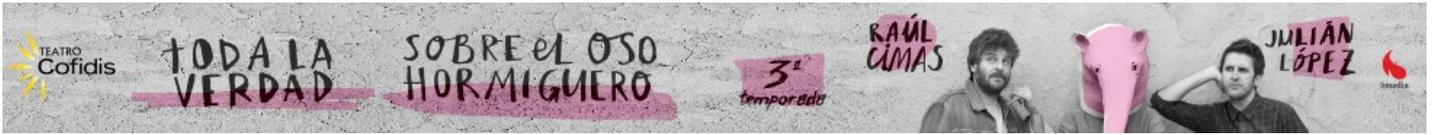
Top 15 delanteros nominados para el mejor equipo del Mundo (Rankings de Futbol)



El Gol Mas más de broma del Mundo! (Rankings de Futbol)

recomendado por

COMENTAR



SOBRE eldiario.es

[QUÉ ES eldiario.es](#) | [EL EQUIPO](#) | [DEFENSOR DE LA COMUNIDAD](#) | [FAQ](#) | [VIVIMOS EN REDES](#)



COLABORA

Nuestra independencia depende de ti. Necesitamos tu apoyo económico para poder hacer un periodismo riguroso y con valores sociales.



HAZTE SOCIO
HAZTE SOCIA
DE eldiario.es

[MÁS INFO](#) | [CREATIVE COMMONS](#) | [AVISO LEGAL Y CONDICIONES DE USO](#) | [CONTACTO](#)