

# NAVEGANDO A TRAVÉS DEL DILUVIO DE DATOS ASTRONÓMICOS

La Astronomía se ha transformado rápidamente en una ciencia de grandes volúmenes de datos. Giga, Tera y próximamente Petabytes deben ser procesados en tiempo real para aprovechar el potencial de los telescopios de nueva generación. Esto pone nuevos desafíos informáticos sobre la mesa: distribución y descripción de datos, taxonomías, semántica, minería de datos, visualización y procesamiento estadístico, dentro de otros. Para tener éxito al abordar estos problemas es indispensable la colaboración entre equipos interdisciplinarios de astrónomos, estadísticos, ingenieros y científicos del área de Computación.



## GUILLERMO CABRERA

Candidato a Doctor en Ciencias de la Computación, Magíster en Ciencias de la Computación, Ingeniero Civil en Computación y Licenciado en Astronomía, Universidad de Chile. Fundador del Laboratorio de Astroinformática del Centro de Modelamiento Matemático de la U. de Chile.

[gcabrera@dim.uchile.cl](mailto:gcabrera@dim.uchile.cl)

Una gran revolución científica se vivió cuando los primeros telescopios se inventaron a principios del siglo XVII. La humanidad comenzó a ver cosas que antes no podía, tales como cráteres en la luna, cometas, los anillos de Saturno e incluso sus lunas. Una nueva revolución se vivió con la creación de placas fotográficas a finales del siglo XIX con las cuales los científicos podían almacenar las observaciones para su futuro análisis. Luego, siguió la revolución digital, con la creación de los llamados Charged-Coupled Devices (CCDs) con los cuales se crearon las primeras cámaras digitales. El astrónomo entonces podía analizar computacionalmente sus imágenes de manera automática (o semi). Hoy en día, vivimos una nueva revolución: la revolución de los datos. Los telescopios están creando cada año más y más datos siguiendo un crecimiento exponencial. Este inmenso volumen de datos está empujando a diversas disciplinas hacia la frontera de sus conocimientos, incluyendo la astronomía, la ingeniería, las estadísticas y la Ciencias de la Computación, entre otras.

Jim Gray vio venir este diluvio años atrás, y lo llamo "el cuarto paradigma" [Tony Hey, 2009]. Según Gray, históricamente existen tres paradigmas científicos clásicos: ciencia empírica (descripción de fenómenos naturales), ciencia teórica (modelos y generalizaciones), y ciencia computacional (complejas simulaciones computacionales). Hoy nos vemos enfrentados a un nuevo paradigma: la ciencia de la exploración de datos (eScience), donde se unifica la teoría con los experimentos y simulaciones en torno al análisis masivo de datos. Esto ha ocurrido en diversas áreas de investigación, donde uno de los ejemplos más claros puede ser encontrado

en la bioinformática (por ejemplo, en genómica) en la cual se producen y analizan terabytes de datos. La Astronomía ha llegado a estos niveles de producción de datos en los últimos diez años, llegando a hablar de la necesidad de una nueva disciplina: la llamada Astroinformática [Borne, et al., 2009].

## GRANDES TELESCOPIOS Y GRANDES VOLÚMENES DE DATOS

Existen distintos telescopios para distintos objetivos científicos. Una primera distinción se puede hacer en función del tipo de radiación electromagnética que se desea observar. Podemos definir esta radiación por el tamaño de su longitud de onda (o, equivalentemente, su frecuencia). De esta forma tenemos telescopios como el Very Large Array (VLA), el Atacama Large Millimeter/submillimeter Array (ALMA) (**Imagen 1**) o el Wilkinson Microwave Anisotropy Probe (WMAP) en el radio (longitud de onda mayor que  $10^3$  m.), el Spitzer Space Telescope (SST) y el James Webb Space Telescope (JWST) en el infrarrojo ( $10^3$ – $10^6$  m.), el Hubble Space Telescope (HST), el Sloan Digital Sky Survey (SDSS), el Large Synoptic Survey Telescope (LSST) (**Imagen 2**) y GAIA en el visible (lo que podemos ver con nuestros ojos,







**IMAGEN 1.**  
EL ATACAMA LARGE MILLIMETER/SUBMILLIMETER ARRAY (ALMA). CADA PAR DE ANTENAS FORMAN UNA LINEA DE BASE, I.E. UN PUNTO EN EL PLANO DE FOURIER. WWW.ALMAOBSERVATORY.ORG  
CRÉDITO: EFE / ARIEL MARINKOVIC.



**IMAGEN 2.**  
FOTOGRAFÍA DEL SITIO E IMAGEN GENERADA POR COMPUTADOR DEL LSST. EL LSST TRABAJARÁ SOBRE VARIABLES ESPACIALES Y TEMPORALES Y PRODUCIRÁ APROXIMADAMENTE 30 TB DE DATOS POR NOCHE A PARTIR DEL 2022. FUENTE: WWW.LSST.ORG

aproximadamente entre 400 y 700 nm.), el Galaxy Evolution Explorer (GALEX) y Hisaki en el ultravioleta ( $\sim 10^{-6}$  –  $10^{-8}$  m.), el Nuclear Spectroscopic Telescope Array (NuSTAR) para rayos-x ( $\sim 10^{-8}$  –  $10^{-11}$  m.), y el Fermi Gamma-ray Space Telescope para rayos gamma (longitud de onda menor que  $< 10^{-11}$  m.) dentro de muchos otros. Aún cuando para observar el universo en distintas longitudes de onda se requiere de distintos telescopios, todos los nuevos grandes instrumentos tienen un problema en común: la gran cantidad de datos a analizar. Por ejemplo, ALMA produce aproximadamente 250 TB por año, el LSST producirá 30 TB por noche (a partir del 2022) y el Square Kilometer Array (**Imagen 3**) producirá más de 100 GB por segundo (2030). En este último caso, la cantidad de datos crudos será tan grande, que no será posible almacenarla, por lo que se deberá procesar mediante hardware y guardar solamente los datos procesados.

Con este gran volumen de datos vienen nuevos problemas, los cuales están delineados en [Borne, et al., 2009]: organización (cómo y dónde), descripción (metadatos), taxonomías o reglas, definición de conceptos y relaciones (semántica), minería de datos y aprendizaje de máquinas, visualización y astroestadística. Todos estos aspectos son los que aborda el área de la Astroinformática.

## PROCESANDO GRANDES VOLÚMENES DE DATOS ASTRONÓMICOS

### PRODUCIENDO IMÁGENES

Uno puede pensar en un telescopio digital clásico de forma similar que en la cámara digital de nuestro celular: un montón de píxeles. Una diferencia importante está en las condiciones de lo que queremos fotografiar. Al tomar una fotografía con nuestro celular a un paisaje o a nuestro rostro favorito, con una iluminación adecuada no tendremos problemas de ruido o de resolución más allá del tamaño de los píxeles. En las imágenes astronómicas la cosa se complica un poco. Los objetos están realmente lejos y antes de ser detectada en el CCD la luz sufre aberraciones al pasar por el telescopio o por la atmósfera (en el caso de telescopios terrestres). Debido a que los CCDs “cuentan” foto-

nes las imágenes contienen ruido Poissoniano además de ruido térmico de distribución Gaussiana. Al mismo tiempo, si queremos llegar realmente profundo debemos tomar imágenes con un tiempo de exposición prolongado, pero lo más corto posible de tal forma de poder observar mayor cantidad de zonas del cielo. Además, al fotografiar (u observar) grandes campos del cielo, hay que tener cuidado con objetos muy brillantes, porque los CCDs se saturan comportándose no lineales (i.e. las “cuentas” no van linealmente con el número de fotones). Por otro lado, se debe también considerar la luz del cielo debido a ciudades u objetos brillantes cercanos (background), los electrones introducidos por segundo por píxel durante la exposición de los CCDs (corriente oscura) y la variación en la eficiencia de cada píxel (flats).

### UTILIZANDO VARIOS TELESCOPIOS COMO UN SOLO GRAN INSTRUMENTO

La resolución de un telescopio se define como la distancia angular mínima a la cual es posible distinguir dos fuentes puntuales como independientes. Esta resolución es directamente proporcional al tamaño del telescopio, e inver-



**IMAGEN 3.** IMAGEN GENERADA POR COMPUTADOR DEL SKA. EL SKA ABARCARÁ UN DIÁMETRO DE APROXIMADAMENTE 3.000 KM Y PRODUCIRÁ APROXIMADAMENTE 100 GB POR SEGUNDO A PARTIR DEL AÑO 2030. FUENTE: WWW.SKATELESCOPE.ORG.

samente proporcional a la longitud de onda de la radiación que se quiere observar. En ese sentido, mientras más grande el lente de un telescopio, mejor resolución tendrá. Al mismo tiempo, mientras más grande sea la longitud de onda de la luz a observar, peor será la resolución. Por supuesto, existen límites físicos para la creación de lentes o espejos (sería un gran desafío crear un lente de 1km. de diámetro, por ejemplo), por lo que la observación de grandes longitudes de onda (como el radio) a alta resolución no es un tema trivial. Para resolver este problema existe la *interferometría*, técnica a través de la cual se puede obtener una imagen de alta resolución mediante la utilización de muchos pequeños telescopios. En ese sentido, al ubicar dos telescopios a una distancia  $D$  obtenemos una resolución similar a la de un telescopio con un lente o espejo de diámetro  $D$ . Ahora sí podemos tener nuestro telescopio de 1 km. de diámetro, pero esto supone nuevos problemas.

En el caso de la interferometría, la señal de cada par de telescopio (o antenas) se correlaciona obteniendo un punto en el plano de Fourier de la imagen, llamado *visibilidad*. En particular, ALMA utiliza un supercomputador diseñado especialmente para realizar las correlaciones entre antenas [Eskoffier, et al., 2007]. A medida que se van agregando más antenas, el plano de Fourier se va poblando (para  $N$  antenas tenemos  $N*(N-1)$  líneas de base considerando las dos orientaciones de cada par). La configuración de antenas puede ir cambiando gracias a la rotación de la tierra o al movimiento físico de ellas, pero los puntos en el plano de Fourier están distribuidos de manera no uniforme. Debido a esto, la reconstrucción de imágenes a partir de las llamadas visibilidades (puntos en el plano de Fourier) no es tan fácil como simplemente realizar una transformada inversa. Para esto existen diversos algoritmos tales como CLEAN [Högbom, 1974], MEM [Cornwell

& Evans, 1985]y métodos Bayesianos [Pina & Puetter, 1993][Sutton & Wandelt, 2006][Cabrera, Casassus & Hitschfeld, 2008]. CLEAN consiste en modelar iterativamente la señal como fuentes puntuales, las cuales son ubicadas en los píxeles de mayor brillo. Para cada una de estas fuentes puntuales se determinan sus visibilidades modeladas y se restan a las observadas, una a una. El Maximum Entropy Method (MEM) es un método variacional donde se optimiza la verosimilitud más un término denominado *entropía*, el cual se determina mediante prueba y error. Los métodos Bayesianos extienden este enfoque reemplazando la entropía por un prior calculado probabilísticamente. Curiosamente, para el caso particular de ALMA y el VLT el paquete oficial de reducción de datos, llamado CASA (Common Astronomy Software Applications) [McMullin, Waters, Schiebel, Young, & Golap, 2007] solamente tiene implementado el algoritmo CLEAN, el cual data de 1974. Esto se debe a razones históricas: en general si un algoritmo es bien comprendido por la comunidad astronómica y funciona relativamente bien, pocos esfuerzos se hacen por mejorarlo. Esto por lo menos hasta hace algunos años, cuando comenzó la real revolución de datos.

## ENFOQUE CLÁSICO VERSUS ASTRONOMÍA DE SURVEYS

El paradigma clásico de observación astronómica es el siguiente: el astrónomo pide tiempo de observación en un telescopio presentando un proyecto. Este proyecto es evaluado por el Time Allocation Comité (TAC), quienes, en caso de aceptar, agendan sus observaciones para alguna fecha particular. Entonces el astrónomo debe ir al telescopio (o envía a alguien) y observar durante una o varias noches. Luego se lleva los datos a la oficina y puede pasar meses (o años) analizando los datos para luego obtener alguna conclusión científica.



Con la creación de los primeros surveys este paradigma ha estado cambiando durante los últimos años. Un survey es un conjunto de datos creados con un telescopio dedicado a observar varias regiones del cielo (o varios objetos) durante un largo período de tiempo. Luego, estos datos se hacen públicos (a veces solo dentro de la colaboración) en forma de imágenes, espectros y sobre todo catálogos a través de grandes bases de datos. En la **Imagen 4**, por ejemplo, podemos ver una imagen obtenida del SDSS. El astrónomo entonces ya no debe ir a observar: puede buscar sus objetos de interés dentro de estas grandes bases de datos, e incluso utilizar todos estos datos para obtener conclusiones científicas a través de herramientas estadísticas.

Aún cuando los surveys existen desde la época de las placas fotográficas (por ejemplo, POSS [Minkowski & Abell, 1963]), la real revolución se dio con los primeros surveys digitales. El Sloan Digital Sky Survey (SDSS) es un ejemplo de esto. El último Data Release del SDSS contiene aproximadamente 70 TB de datos, donde la mitad son datos crudos y la otra mitad son catálogos [SDSS Collaboration, 2013]. Al igual que los telescopios, existen distintos surveys para distintos objetivos científicos. La **Tabla 1** muestra distintos surveys y su volumen de datos. Durante los últimos años se han creado proyectos que permitirán incluir también la variable de tiempo a las observaciones. Dos ejemplos de esto son el Dark Energy Survey (DES), el cual comenzó a operar en 2013, y el Large Synoptic Survey Telescope (LSST) [Ivezic, et al., 2008], el cual se espera que comience a operar en 2022. Este último, tomará una imagen completa del cielo Sur cada tres noches. En otras palabras, tendremos una película de todo el cielo con un *timeframe* de tres noches. El LSST producirá diariamente 30 TB de datos (15 TB crudos más 15 TB procesados), los cuales deben ser procesados completamente en menos de un día, ya que a la siguiente noche tendremos nuevamente 30 TB más.

Además de la reducción inicial de las imágenes descrita anteriormente (también llamada pri-



**IMAGEN 4.** TIPO DE IMÁGENES A SER ANALIZADAS POR LOS SURVEYS. SOBRE ÉSTAS IMÁGENES ES NECESARIO REALIZAR DETECCIÓN, ASTROMETRÍA, FOTOMETRÍA Y CLASIFICACIÓN AUTOMÁTICAMENTE. FUENTE: SLOAN DIGITAL SKY SURVEY. WWW.SDSS.ORG

mera capa), los surveys deben realizar un conjunto de tareas propias de sus tipos de datos incluyendo: detección automática de objetos (estrellas, galaxias, asteroides, etc.) [Stetson, 1987] [Bijaoui & Rué, 1995] [Bertin & Arnouts, 1996] [Miller, et al., 2001], astrometría (calzar distintas imágenes en un mismo sistema de coordenadas)[Valdés, Campusano, Velásquez, & Stetson, 1995] [Lang, Hogg, Mierle, Blanton, & Roweis, 2010], calce de resolución entre imágenes (PSF matching) [Bertin, 2011], fotometría (calcular exactamente cuanta luz emite cada objeto) [Bertin & Arnouts, 1996] y catalogar las distintas fuentes (ajustes de modelos, clasificación, series de tiempo, etc., ver por ejemplo [Ivezic, et al., 2008]). A todo esto se le suele llamar segunda capa y los encargados de estas tareas son generalmente los mismos observatorios.

La tercera y última capa incluye el análisis astroestadístico de los catálogos y la obtención de conclusiones científicas. Esto no necesariamente lo realizan los mismos observatorios, sino que generalmente es aquí donde el astrónomo de la universidad se conecta a la base de datos para acceder ya sea a catálogos o imágenes y procesarlos. Es aquí además donde se requiere

de experiencia en el área de Minería de Datos y Estadística, por lo que los equipos interdisciplinarios juegan un rol fundamental.

## EQUIPOS INTERDISCIPLINARIOS PARA PROBLEMAS INTERDISCIPLINARIOS

Ya lo decía Jim Gray antes de desaparecer misteriosamente: los problemas científicos de grandes volúmenes de datos requieren de equipos interdisciplinarios. En nuestro caso, es impensable que un equipo compuesto 100% por astrónomos sea capaz de enfrentar todos los problemas anteriormente descritos. Gray decía que para que un proyecto en eScience sea exitoso se requiere de cuatro piezas fundamentales: los científicos, quienes proporcionan las preguntas a ser respondidas, los plomeros quienes son los encargados de diseñar y mantener las bases de datos, los mineros quienes desarrollan los algoritmos de minería de datos, y los desarrolladores de las herramientas de visualización de preguntas y respuestas. La comunicación entre integrantes de estos equipos debe ser fluida, por lo que,

Survey	Año de inicio	Año de fin	Longitud de Onda	Ubicación	Tamaño total de datos
POSS	1949	1958	visible	California, USA	3TB
2MASS	1997	2001	infrarojo	Arizona, USA+Chile	10TB
GALEX	2003	2013	ultravioleta	Espacio	30TB
SDSS	2000	2020	visible	New Mexico, USA	70TB
GAIA	2013	2020	visible	Espacio	1PB
DES	2013	2018	visible	Chile	1-5PB
PanSTARRS	2008	-	visible	Hawaii	40PB
LSST	2022	2032	visible	Chile	75PB

**TABLA 1.**  
SURVEYS ASTRONÓMICOS.

aún cuando no es necesario que todos sean expertos en todas las áreas, si es necesario que conozcan al menos superficialmente un poco de cada tema. De esta forma, los expertos en computación y estadística deben saber algo de astronomía y los astrónomos deben entender al menos los aspectos básicos de computación y estadística.

Durante los últimos años se han creado varios de estos grupos interdisciplinarios, los cuales han sido exitosos en el desarrollo de nuevos métodos para abordar problemas astronómicos a través de grandes volúmenes de datos. Algunos ejemplos son: detección de transientes [Bailey, Aragon, Romano, Thomas, Weaver, & Wong, 2007] [Brink, et al., 2013], clasificación morfológica de galaxias [Ball, et al., 2004] [Hurtas-Company, Aguerri, Bernardi, Mei, & Sánchez Almeida, 2011] [Lintott, et al., 2011] clasificación de espectros [Daniel, Connolly, Schneider, Vanderplas, & Xiong, 2011], obtención de períodos en curvas de luz [Huijse, Estévez, Zegers, Principe, & Protopapas, 2011] [Graham, et al., 2013], y descubrimiento de relaciones entre variables

astronómicas [Graham, Djorgovski, Mahabal, Donalek, & Drake, 2013] entre otros.

Otro aspecto interesante de estas colaboraciones, es cómo al enfrentar problemas astronómicos se descubren nuevos problemas computacionales y se crean nuevos algoritmos, los cuales incluso pueden ser aplicados a otras áreas. Algunos ejemplos de esto han ocurrido al indexar y calzar series de tiempo [Keogh, Wei, Xi, Vlachos, Lee, & Protopapas, 2009], en clasificación con datos incompletos [Pichara & Protopapas, 2013] y detección y corrección de etiquetas sesgadas en aprendizaje supervisado [Cabrera, Miller & Schneider, 2014].

En el caso particular de Chile, en el Centro de Modelamiento Matemático de la Universidad de Chile llevamos más de cinco años trabajando en estos temas. Hemos creado un laboratorio en Astroinformática (AstroLab) formado por un gran equipo interdisciplinario que incluye científicos del área de astronomía, modelamiento matemático, estadística, machine learning, HPC, procesamiento de imágenes, ingeniería

eléctrica, etc. Algunos de los proyectos que se trabajan actualmente en el AstroLab incluyen detección de transientes en tiempo real, reconstrucción de imágenes interferométricas, clasificación morfológica de galaxias, reglas de asociación para líneas moleculares, registro y superresolución de imágenes astronómicas, entre otros.

## CONCLUSIONES

CON LOS NUEVOS INSTRUMENTOS, DURANTE LOS PRÓXIMOS AÑOS SE NOS VIENEN ENCIMA GIGA, TERA Y PETABYTES DE DATOS ASTRONÓMICOS LOS CUALES DEBEN SER PROCESADOS EN TIEMPO REAL. ESTOS DATOS SERÁN COMPLETAMENTE NUEVOS Y LO QUE ELLOS CONTENGAN NO LO SABEMOS AÚN. ESPERAMOS ENCONTRAR NUEVOS TIPOS DE OBJETOS Y NUEVA ASTROFÍSICA, ASÍ COMO TAMBIÉN EXPLICAR LOS MISTERIOS QUE AÚN NO HEMOS PODIDO RESOLVER. PARA TENER ÉXITO EN ESTE DESAFÍO ES DE SUMA IMPORTANCIA LA CREACIÓN DE EQUIPOS INTERDISCIPLINARIOS DE INVESTIGADORES PARA ASÍ PODER ATACAR EL PROBLEMA CON TODO TIPO DE HERRAMIENTAS, TANTO MATEMÁTICAS Y ESTADÍSTICAS COMO COMPUTACIONALES. ■

## AGRADECIMIENTOS

Muchas gracias a Sebastián Pérez por ayudar a recopilar algunos de los datos de este artículo. Este trabajo ha sido parcialmente financiado por el Centro de Modelamiento Matemático de la Universidad de Chile, por el proyecto Fondecy D1111060, y por la Beca de Doctorado Nacional de CONICYT.

## BIBLIOGRAFÍA

- Cabrera, G. F., Casassus, S., & Hirschfeld, N. (2008). Bayesian Image Reconstruction Based on Voronoi Diagrams. *The Astrophysical Journal*, 672, 1272-1285.
- Cabrera, G. F., Miller, C. J., & Schneider, J. (2014). Systematic Labeling Bias: De-Biasing Where Everyone Is Wrong. 22nd International Conference on Pattern Recognition. Stockholm.
- Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. (2010). Astrometry.net: Blind Astrometric Calibration of Arbitrary Astronomical Images. *The Astronomical Journal*, 139, 1782-1800.
- Lintott, C., Schawinski, K., Bamford, S., Slosar, A., Land, K., Thomas, D., et al. (2011). Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 410, 166-178.
- Cornwell, T. J., & Evans, K. F. (1985). A simple maximum entropy deconvolution algorithm. *Astronomy and Astrophysics*, 143, 77-83.
- Ball, N. M., Loveday, J., Fukugita, M., Nakamura, O., Okamura, S., Brinkmann, J., et al. (2004). Galaxy types in the Sloan Digital Sky Survey using supervised artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, 348, 1038-1046.
- Bailey, S., Aragon, C., Romano, R., Thomas, R. C., Weaver, B. A., & Wong, D. (2007). How to Find More Supernovae with Less Work: Object Classification Techniques for Difference Imaging. *The Astrophysical Journal*, 665, 1246-1253.
- Bertin, E. (2011). Automated Morphometry with SExtractor and PSFEx. *Astronomical Data Analysis Software and Systems XX* (p. 435). Boston: Evans, I. N.; Accomazzi, A.; Mink D.J.; Rots A.H.
- Bertin, E., & Arnouts, S. (1996). SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement*, 117, 393-404.
- Bijaoui, A., & Rué, F. (1995). A multiscale vision model adapted to the astronomical images. *Signal processing*, 46 (3), 345-362.
- Borne, K., Accomazzi, A., Bloom, J., Brunner, R., Burke, D., Butler, N., et al. (2009). *Astroinformatics: A 21st Century Approach to Astronomy. Astro2010: The Astronomy and Astrophysics Decadal Survey*, Position Papers, no. 6.
- Brink, H., Richards, J. W., Poznanski, D., Bloom, J. S., Rice, J., Negahban, S., et al. (2013). Using machine learning for discovery in synoptic survey imaging data. *Monthly Notices of the Royal Astronomical Society*, 435, 1047-1060.
- Daniel, S. F., Connolly, A., Schneider, J., Vanderplas, J., & Xiong, L. (2011). Classification of Stellar Spectra with Local Linear Embedding. *The Astronomical Journal*, 142, 203.

- Escoffier, R. P., Comoretto, G., Webber, J. C., Baudry, A., Broadwell, C. M., Greenberg, J. H., et al. (2007). The ALMA correlator. *Astronomy and Astrophysics*, 462 (2), 801-810.
- Graham, M. J., Djorgovski, S. G., Mahabal, A. A., Donalek, C., & Drake, A. J. (2013). Machine-assisted discovery of relationships in astronomy. *Monthly Notices of the Royal Astronomical Society*, 431, 2371-2384.
- Graham, M. J., Drake, A. J., Djorgovski, S. G., Mahabal, A. A., Donalek, C., Duan, V., et al. (2013). A comparison of period finding algorithms. *Monthly Notices of the Royal Astronomical Society*, 434, 3423-344.
- Högbom, J. A. (1974). Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy & Astrophysics Supplement Series*, 15, 417 - 426.
- Huertas-Company, M., Aguerri, J. A., Bernardi, M., Mei, S., & Sánchez Almeida, J. (2011). Revisiting the Hubble sequence in the SDSS DR7 spectroscopic sample: a publicly available Bayesian automated classification. *Astronomy and Astrophysics*, 525, A157.
- Huijse, P., Estévez, P. A., Zegers, P., Príncipe, J. C., & Protopapas, P. (2011). Period Estimation in Astronomical Time Series Using Slotted Correntropy. *IEEE Signal Processing Letters*, 18, 371-374.
- Ivezic, Z., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., ALSayyad, Y., et al. (May de 2008). LSST: from Science Drivers to Reference Design and Anticipated Data Products. arxiv.
- Keogh, E., Wei, L., Xi, X., Vlachos, M., Lee, S.-H., & Protopapas, P. (2009). Supporting exact indexing of arbitrarily rotated shapes and periodic time series under Euclidean and warping distance measures. *The VLDB Journal*, 18, 611-630.
- McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. (2007). CASA Architecture and Applications. In R. A. Shaw, F. Hill, & D. J. Bell (Ed.), *Astronomical Data Analysis Software and Systems XVI*, 376, p. 127. Tucson.
- Miller, C. J., Genovese, C., Nichol, R. C., Wasserman, L., Connolly, A., Reichart, D., et al. (2001). Controlling the False-Discovery Rate in Astrophysical Data Analysis. *The Astronomical Journal*, 122, 3492-3505.
- Minkowski, R. L., & Abell, G. O. (1963). The National Geographic Society-Palomar Observatory Sky Survey. In K. A. Strand, *Basic Astronomical Data: Stars and stellar systems* (p. 481). Chicago: University of Chicago Press.
- Pichara, K., & Protopapas, P. (2013). Automatic Classification of Variable Stars in Catalogs with Missing Data. *The Astrophysical Journal*, 777, 83.
- Pina, R. K., & Puetter, R. C. (1993). Bayesian image reconstruction - The pixon and optimal image modeling. *Astronomical Society of the Pacific, Publications*, 105, 630-637.
- SDSS Collaboration. (2013). Data Volume Table. Retrieved 09 de 2014 from SDSS III: [https://www.sdss3.org/dr10/data\\_access/volume.php](https://www.sdss3.org/dr10/data_access/volume.php)
- Stetson, P. B. (1987). DAOPHOT - A computer program for crowded-field stellar photometry. *Astronomical Society of the Pacific, Publications*, 99, 191-222.
- Sutton, E. C., & Wandelt, B. D. (2006). Optimal Image Reconstruction in Radio Interferometry. *The Astrophysical Journal Supplement Series*, 162, 401-416.
- Tony Hey, S. T. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.
- Valdés, F. G., Campusano, L. E., Velásquez, J. D., & Stetson, P. B. (1995). FOCAS Automatic Catalog Matching Algorithms. *Publications of the Astronomical Society of the Pacific*, 107, 1119.