

Precisa-se de um título desesperadamente

André Luiz de Amorim

Orientador:

Prof. Dr. Roberto Cid Fernandes Jr.

• • •

Universidade Federal de Santa Catarina
Centro de Ciências Físicas e Matemáticas
Departamento de Física

Dissertação de mestrado apresentada ao Departamento de Física da UFSC em preenchimento parcial dos requisitos para obtenção do título de Mestre em Física.

Trabalho financiado pela Capes.

Florianópolis (SC) – 24 de outubro de 2011

Dedica

Agradecimientos

Agradece

Resumo

Resumo.

Abstract

Abstract.

Sumário

1	Introdução	1
1.1	A nova era da astronomia	1
1.2	<i>STARLIGHT</i> + <i>SDSS</i>	1
1.3	<i>GALEX</i>	1
1.4	Trabalhos Anteriores	1
1.5	Este Trabalho	1
1.6	Organização deste trabalho	2
2	O <i>Galaxy Evolution Explorer</i> (<i>GALEX</i>)	3
2.1	Objetivos do <i>GALEX</i>	3
2.2	Histórico do estudo do céu no ultravioleta	6
2.3	Resultados obtidos pelo <i>GALEX</i>	7
2.4	Data releases e banco de dados	10
3	Crossmatch entre <i>SDSS/STARLIGHT</i> e <i>GALEX</i>	12
3.1	Banco de dados do <i>SDSS</i>	12
3.1.1	Migração de OODBMS para RDBMS	13
3.1.2	<i>SkyServer</i>	13
3.1.3	<i>CasJobs</i>	14

3.2	Banco de dados do <i>STARLIGHT</i>	16
3.2.1	Importação para o RDBMS	16
3.2.2	Estrutura do banco de dados	17
3.2.3	Amostra do <i>STARLIGHT</i>	18
3.3	Crossmatch <i>SDSS/GALEX</i>	20
3.3.1	Identificação dos objetos	20
3.3.2	Indexação por <i>Hierarchical Triangular Mesh</i>	21
3.4	Definição das amostras <i>SDSS/STARLIGHT</i> e <i>GALEX</i>	23
3.5	Correções na fotometria UV	25
3.5.1	Poeira	25
3.5.2	K-correct	25
4	Problemas astrofísicos	26
5	Conclusões e Perspectivas	27
5.1	Este trabalho	27
5.2	Trabalhos Futuros	27
A	Anexo 1: Manual de Acesso aos dados do <i>STARLIGHT</i> + Galex	28
	Referências Bibliográficas	I

Lista de Figuras

2.1	Curvas de transmissão dos filtros do <i>GALEX</i>	4
2.2	<i>Footprint</i> dos <i>surveys</i> <i>GALEX</i> AIS, MIS e SDSS	6
2.3	Diagrama cor-magnitude em ultravioleta.	8
2.4	Tela do programa <i>GalexView</i>	11
3.1	Telas do <i>SkyServer</i>	14
3.2	Tela do <i>CasJobs</i>	15
3.3	Esquema do banco de dados do <i>STARLIGHT</i>	18
3.4	Esquema do banco de dados do <i>SDSS</i>	19
3.5	Esquema da tabela de índices da amostra do <i>STARLIGHT</i>	19
3.6	<i>Query</i> para atualizar os índices da amostra de galáxias do <i>STARLIGHT</i>	20
3.7	Todo	22
3.8	Todo	22
3.9	Esquema da tabela de em crossmatch entre objetos do <i>GALEX</i> e do <i>SDSS</i>	23
3.10	<i>Query</i> para o <i>match</i> entre os objetos da amostra do <i>STARLIGHT</i> e <i>GALEX</i> AIS.	24

Lista de Tabelas

2.1	<i>Surveys</i> realizados pelo <i>GALEX</i>	5
2.2	Campos observados em cada <i>General Release</i> do <i>GALEX</i>	10
3.1	Número de objetos correlacionados entre <i>SDSS</i> DR7 e <i>GALEX</i> GR6. . .	21

Capítulo 1

Introdução

1.1 A nova era da astronomia

1.2 STARLIGHT + SDSS

1.3 GALEX

Resumo do Galex, o que é, como funciona, motivação. Por que UV? Ir pra outros comprimentos de onda. Limitações do UV. Propaganda do GALEX. Necessidade de ir para outros λ , e qual ciência pode ser feita com cada faixa.

1.4 Trabalhos Anteriores

Observatórios virtuais. Crossmatch.

1.5 Este Trabalho

Crossmatch entre fontes SDSS do *STARLIGHT* e do Galex. Adicionar alguns problemas astronômicos.

1.6 Organização deste trabalho

Capítulo 2

O Galaxy Evolution Explorer (GALEX)

2.1 Objetivos do GALEX

O *Galaxy Evolution Explorer* (GALEX) é um telescópio espacial de pequeno porte da NASA¹, lançado em 28 de abril de 2003 para conduzir um *survey* de todo o céu numa faixa espectral do ultravioleta (1350–2750Å). O objetivo principal do GALEX é estudar a evolução da taxa de formação estelar em galáxias (Martin et al. 2005). Os dados coletados pela missão são publicados em *Data Releases* periódicos, denominados *General Releases*. Este trabalho foi realizado sobre os dados do sexto *General Release*, GR6.

A missão consiste em uma série de *surveys* fotométricos e espectroscópicos (ver tabela 2.1). Destes, os principais *surveys* são o *All Sky Survey* (AIS) e o *Medium Imaging Survey* (MIS), que foram utilizados neste trabalho. O imageamento é feito em duas bandas espectrais: ultravioleta distante (*far ultraviolet*, FUV), de 1350 a 1750Å, e ultravioleta próximo (*near ultraviolet*, NUV), de 1750 a 2750Å. As curvas de transmissão dos filtros utilizados nessas bandas podem ser visto na figura 2.1. A espectroscopia é feita inserindo-se no caminho ótico um *grism*, que consiste num prisma combinado com uma rede de difração. Obtém-se deste modo um espectro de baixa resolução para cada objeto na imagem, conforme descrito por Morrissey et al. (2007).

Os *surveys* do GALEX foram planejados de forma a se valer de outros *surveys* já existentes em outros comprimentos de onda. A figura 2.2 mostra a sobreposição da

¹NASA *Small Explorer* (SMEX) - <http://explorers.gsfc.nasa.gov/missions.html>

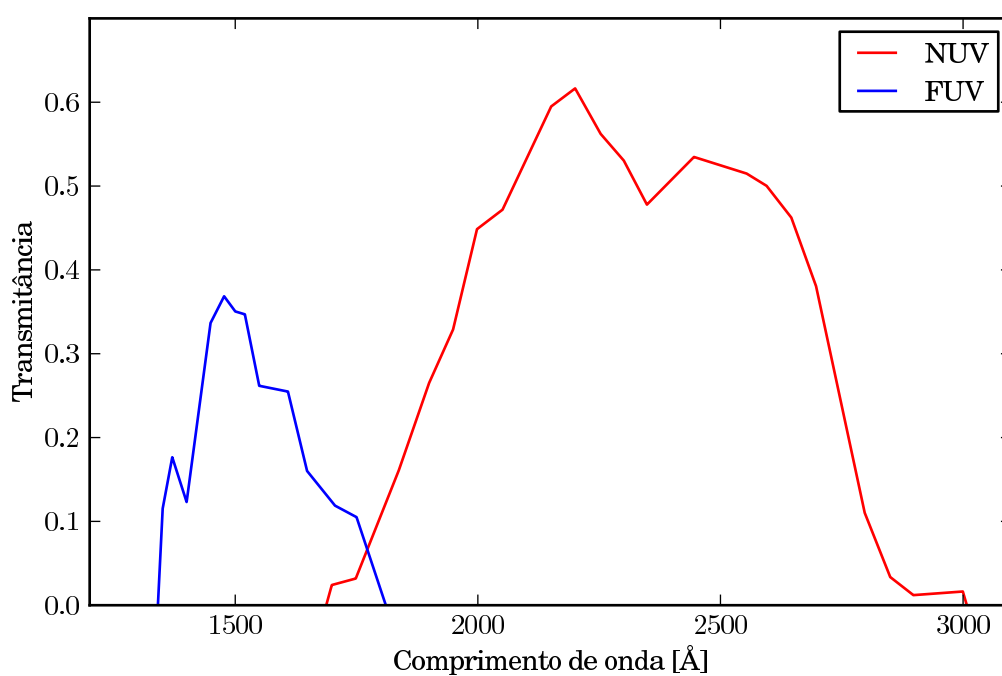


Figura 2.1: Curvas de transmissão dos filtros do *GALEX*, medidas em laboratório (Morrissey et al. 2005). Retirada do *website* mantido por Peter Capak: <http://www.astro.caltech.edu/~capak/cosmos/filters/>

Tabela 2.1: *Surveys* realizados pelo *GALEX*. O CAI consiste em observações de anãs brancas para calibração. A cobertura do céu é dada em graus quadrados. No caso do NGS, a magnitude limite é dada em unidades de densidade superficial de magnitude. Informações retiradas de Martin et al. (2005).

<i>Survey</i>	Cobertura do céu	Mag. AB limite
<i>Calibration Imaging (CAI)</i>	-	-
<i>All-sky Imaging Survey (AIS)</i>	26000	20.5
<i>Medium Imaging Survey (MIS)</i>	1000	23
<i>Deep Imaging Survey (DIS)</i>	80	25
<i>Nearby Galaxy Survey (NGS)</i>	80	27.5
<i>Wide Field Spectroscopic Survey (WSS)</i>	80	20
<i>Medium-deep Spectroscopic Survey (MSS)</i>	8	21.5–23
<i>Deep Spectroscopic Survey (DSS)</i>	2	23–24

área observada ² pelos *surveys* AIS e MIS do *GALEX* e do *Sloan Digital Sky Survey (SDSS)*. O objetivo primário da missão do *GALEX* é calibrar a taxa de formação estelar no universo local e a determinar o histórico cosmológico de formação estelar entre os *redshifts* $0 < z < 2$ (Martin et al. 2005). A comparação com dados de *surveys* em outros comprimentos de onda tem um papel fundamental no cumprimento deste objetivo.

²*Footprint*, no linguajar astronômico.

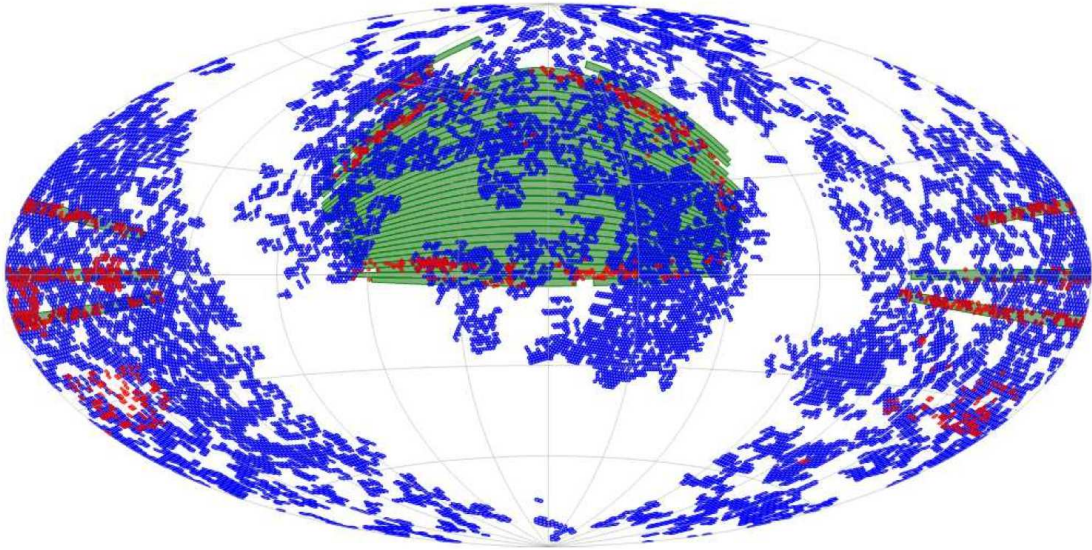


Figura 2.2: *Footprint* dos *surveys* GALEX GR2+3 AIS (azul), MIS (vermelho) e SDSS DR6 (verde), de Budavári et al. (2009)

2.2 Histórico do estudo do céu no ultravioleta

A camada de ozônio, tão desejável pela proteção que oferece aos seres vivos, cobra a sua taxa na astronomia. Observações na banda ultravioleta precisam ser feitas fora da atmosfera terrestre, portanto não é de se estranhar que o trabalho nesta faixa espectral tenha progredido menos do que na faixa do óptico e do infravermelho.^[citation needed]

O primeiro trabalho sistemático de observação UV foi feito pelo *Orbiting Astronomical Observatory 2* (Code et al. 1970), obtendo fotometria e espectroscopia de estrelas brilhantes, aglomerados globulares e galáxias próximas. Durante as décadas de 1970 e 1980, este e outros satélites como o TD-1 (Boksenberg et al. 1973), o *Astronomical Netherlands Satellite* (van Duinen et al. 1975) e o *International Ultraviolet Explorer* (Kondo & Wamsteker 1987) – o primeiro satélite a utilizar um detetor de imageamento UV – forneceram os dados fundamentais para os modelos de síntese de população estelar de galáxias. *Surveys* de campo amplo foram feitos por uma câmera lunar erguida por astronautas da *Apollo 16* (Carruthers 1973), a bordo do *Skylab* (Henize et al. 1975) e pelo instrumento *FAUST* a bordo do *Spacelab* (Bowyer et al. 1993). Muitas imagens UV também foram obtidas pelo *Ultraviolet Imaging Telescope* em duas missões em ônibus espacial (Stecher et al. 1997).

2.3 Resultados obtidos pelo GALEX

o *GALEX* fez o primeiro *survey* do céu inteiro em UV. As regiões próximas ao plano da Galáxia foram evitados para não danificar os detetores. Pode-se ter uma idéia do sucesso desta missão considerando a grande quantidade de artigos publicados³. Abaixo segue um resumo dos resultados mais notáveis.

Wyder et al. (2007) analisam a distribuição de galáxias em função da cor UV e da magnitude absoluta no universo local. Esta distribuição é conhecida como *Diagrama Cor-Magnitude* (CMD, na sigla em inglês para *Color-Magnitude Diagram*). Os autores usam *redshifts* e fotometria óptica obtidas do *SDSS* junto com fotometria UV do *survey* MIS do *GALEX*. A amostra do *SDSS* é correlacionada com a do *GALEX* procurando o objeto do *GALEX* mais próximo de cada objeto *SDSS* até um limite de 4" (4 segundos de arco).

O diagrama cor-magnitude elaborado por Wyder et al. mostra a separação das galáxias nas sequências azul e vermelha (figura 2.3). Esta distribuição bimodal é um resultado bem conhecido na astronomia.^[citation needed] Porém, diferente do diagrama cor-magnitude para a faixa espectral do óptico, a distribuição de cores em UV não pode ser ajustada somente pela soma de duas gaussianas, há um excesso de objetos nas cores intermediárias entre os picos azul e vermelho. A boa separação entre as sequências é atribuída a uma maior sensibilidade à formação estelar recente.

Martin et al. (2007) investigaram as propriedades das galáxias entre as sequências vermelha e azul para a mesma amostra citada acima. As galáxias nesta região intermediária são preferencialmente galáxias com núcleo ativo (*Active Galactic Nucleus*, AGN). Os autores estimam o fluxo de massa de galáxias indo da sequência azul para a vermelha.

Ainda para a mesma amostra, Schiminovich et al. (2007) investigaram a correlação entre a morfologia das galáxias e a sua posição no CMD. A função de luminosidade UV do universo local é medida – pela primeira vez, segundo os autores – com relação aos parâmetros estruturais e à inclinação das galáxias.

A missão do *GALEX* se encerra em 31 de dezembro de 2011. Dados coletados após o GR6, como as observações no mesmo campo utilizado na missão Kepler (?), observações de M31 e da Nuvem de Magalhães, entre outros, serão liberados num último *data release*,

³Há uma lista com as mais de 500 publicações relacionadas ao projeto do *GALEX* em <http://www.galex.caltech.edu/researcher/publications.html>

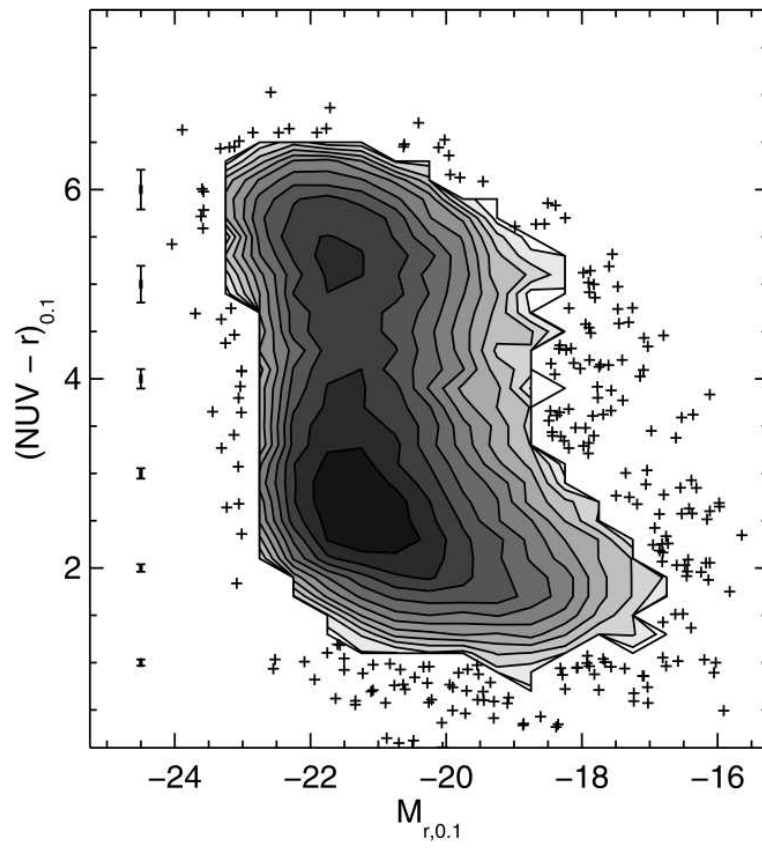


Figura 2.3: Diagrama cor-magnitude em ultravioleta. (Wyder et al. 2007, figura 7).

GR7. Os dados obtidos pelo *GALEX* permanecerão disponíveis publicamente no MAST.

2.4 Data releases e banco de dados

Os dados obtidos pelo *GALEX* são armazenados no *Multi-Mission archive at the Space Telescope Science Institute* (MAST). O acesso a estes dados é público, a liberação é feita anualmente em *General Releases* (GR). Os dados consistem basicamente em imagens e catálogos, divididos em campos (*tiles*) com área de aproximadamente 1,2 graus quadrados. Devido ao modo como o *GALEX* faz as observações, um determinado objeto pode estar presente em mais de um campo. A tabela 2.2 mostra o número cumulativo de campos observados por *survey* em cada GR⁴. Observações de pesquisadores convidados (*Guest Investigators*, GI) foram selecionadas de forma a complementar os *surveys*.

Tabela 2.2: Campos observados em cada *General Release* do *GALEX*.

<i>Release</i>	AIS	DIS	MIS	NGS	GI	CAI	Espectros	Total
GR1	3074	14	112	52	-	-	7	3259
GR2/GR3	15721	165	1017	296	288	20	41	17548
GR4/GR5	28269	292	2161	458	788	38	174	32180
GR6	28889	338	3479	480	1314	51	-	34551

Para facilitar o acesso aos dados do *GALEX*, o MAST desenvolveu uma ferramenta chamada *GalexView*, utilizando tecnologia *Adobe Flex*⁵. Desta forma o *GalexView* pode ser acessado através de seu *website*⁶ em qualquer *web browser* que tenha suporte ao *Adobe Flash Player*⁷.

Através do *GalexView* é possível fazer buscas, visualizar e obter imagens e catálogos dos campos do *GALEX*. As buscas podem ser feitas de forma bastante versátil, tanto pelo nome do objeto quanto pelas coordenadas do céu. O formato de entrada é flexível o suficiente para evitar os problemas causados por idiosincrasias na notação de coordenadas (por exemplo, tanto “14h03m12.6s +54d20m56.7s” quanto “14 03 12.6 54 20 56.7” ou “210.83 54.35” apontam para a mesma região). A sua interface (figura 2.4) permite filtrar

⁴Informações retiradas do em website do GR6: <http://galex.stsci.edu/GR6/>

⁵*Adobe Flex* é um *framework* de código aberto que permite desenvolver aplicações para *web browsers*. Ver <http://www.adobe.com/products/flex.html>.

⁶*GalexView*: <http://galex.stsci.edu/GalexView/>

⁷*Adobe Flash Player* é uma extensão multiplataforma para *web browsers* que provê capacidade de visualização de conteúdo *flash* gerado tanto pelos seus editores proprietários quanto por ferramentas de terceiros. Ver <http://www.adobe.com/products/flashplayer/>.

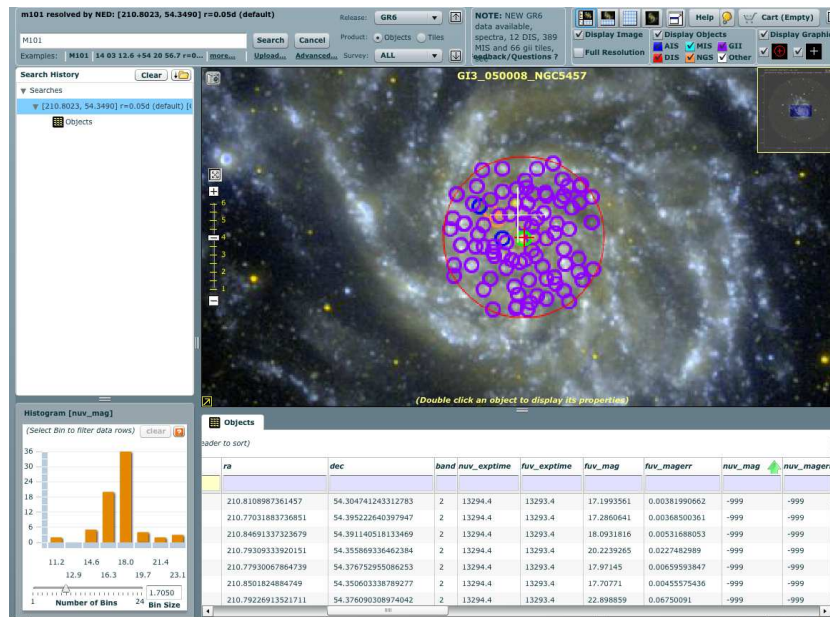


Figura 2.4: Tela do programa *GalaxView*, com a visualização da galáxia M101.

o conteúdo retornado pelas buscas, separando por *surveys*. Há também uma ferramenta de histograma, permitindo filtrar pelos valores das colunas dos catálogos. Os objetos selecionados na busca aparecem marcados na visualização da imagem. Utilizando um sistema do tipo “carrinho de compras”, pode-se selecionar campos e objetos de interesse, para ao final do uso do sistema baixar toda a seleção de uma vez.

Tanto o *GalaxView* quanto outras ferramentas de busca do MAST, como o *GALEX Search Form* e o *GALEX Tilelist*, são construídos sobre um *banco de dados relacional* acessado através da linguagem *SQL* (Chamberlin & Boyce 1974). Muito comum na indústria, bancos de dados relacionais dispõem em geral de uma vasta gama de ferramentas para gerenciamento dos dados. Uma de suas grandes vantagens é o uso de índices⁸ para agilizar o acesso a dados. Embora a tecnologia exista desde a década de 1970 (Codd 1970), até uma década atrás suas vantagens eram praticamente negligenciadas na astronomia.

Bancos de dados relacionais e ferramentas para gerenciamento e acesso a dados serão tratados com mais detalhes no capítulo 3.

⁸Um índice numa tabela de banco de dados é uma estrutura que copia partes da tabela numa determinada ordem, de forma a aumentar a velocidade de acesso aos dados ao custo de espaço de armazenamento.

Capítulo 3

Crossmatch entre SDSS/STARLIGHT e GALEX

Tradicionalmente astrônomos armazenam seus dados em arquivos texto ou binários contendo um registro por linha, de uma forma tecnicamente conhecida como *flat file*. Buscas neste tipo de banco de dados são feitas examinando individualmente cada registro do arquivo. Com o volume de dados obtido pelo *GALEX* (aproximadamente 222 milhões de objetos, 34 mil campos)^[citation needed], o uso de arquivos simples para armazenamento de dados se torna inviável.^[citation needed] É preciso “profissionalizar” o gerenciamento de dados de um *survey* desta escala.

3.1 Banco de dados do SDSS

Um dos maiores responsáveis pela promoção do uso de bancos de dados relacionais na astronomia é o projeto *Sloan Digital Sky Survey* (*SDSS*). Inicialmente o *SDSS* utilizou um *sistema de gerenciamento de banco de dados orientado a objetos* (Maier et al. 1986) (OODBMS, na sigla em inglês). Após pouco mais de um ano a abordagem se mostrou inadequada: entre os principais problemas, uma linguagem de *query* inadequada e performance ruim. O motivo, segundo Thakar et al. (2004), foi a incapacidade da empresa desenvolvedora do OODBMS em prover novas funcionalidades requisitadas pelo projeto e correção de *bugs*, bem como em acompanhar o crescimento da performance do *hardware*.

3.1.1 Migração de OODBMS para RDBMS

Todo o banco de dados do *SDSS* foi migrado para um *sistema de gerenciamento de banco de dados relacional* (Codd 1970) (RDBMS, na sigla em inglês) num esforço guiado por Thakar et al.. RDBMS pode ser considerado o padrão da indústria. Praticamente todas as linguagens de programação tem bibliotecas de interface às implementações de RDBMS comerciais mais comuns (Oracle, IBM e Microsoft, Postgres). Há uma diversidade de ferramentas para desenvolvimento e gerenciamento de RDBMS. E talvez o maior benefício de todos, o acesso aos dados é feito utilizando uma linguagem padronizada: *Simple Query Language*, ou simplesmente SQL (Chamberlin & Boyce 1974). A migração dos dados do *SDSS* para um RDBMS comercial implicou num aumento significativo da performance do acesso aos dados, e resultou no desenvolvimento do *SkyServer*¹. O servidor de banco de dados escolhido pelo *SDSS* foi o *Microsoft SQL Server*.

A comparação entre OODBMS e RDBMS no caso particular do *SDSS* não implica necessariamente a superioridade do segundo em relação ao primeiro. Tanto a abordagem orientada a objetos quanto a abordagem relacional tem suas vantagens e desvantagens. O estudo de caso do *SDSS* é apenas uma evidência anedótica em favor do uso de bancos de dados relacionais. No entanto, para aplicações semelhantes ao *SDSS*— *surveys* astronômicos com volumes imensos de dados — vale a pena apostar no sucesso dos RDBMS.

3.1.2 SkyServer

O *SkyServer* é um *website* (figura 3.1) que provê acesso aos dados armazenados no banco de dados do *SDSS* (Szalay et al. 2002). O acesso mais simples pode ser feito através de um atlas de locais famosos (*famous places*), que mostra imagens coloridas de objetos celestes conhecidos. Há formulários para buscas mais sérias, gerando coleções de imagens, espectros e tabelas de dados. No *SkyServer* é possível fazer buscas avançadas utilizando SQL, embora haja limites de tempo de execução e de quantidade de objetos retornados. Esta limitação é contornada através do sistema *CasJobs*, que é tratado na seção 3.1.3.

É importante ressaltar que é possível (de fato, a equipe do *SDSS* encoraja) criar *mirrors*² do *SkyServer*. Tanto o banco de dados do *SDSS* quanto o código fonte do *SkyServer* estão disponível no próprio *website* do *SkyServer*. Há um clone do banco de

¹*SDSS SkyServer*: <http://skyserver.sdss.org/>

²*Mirror*: Espelho, em inglês. Clone de um *website*.

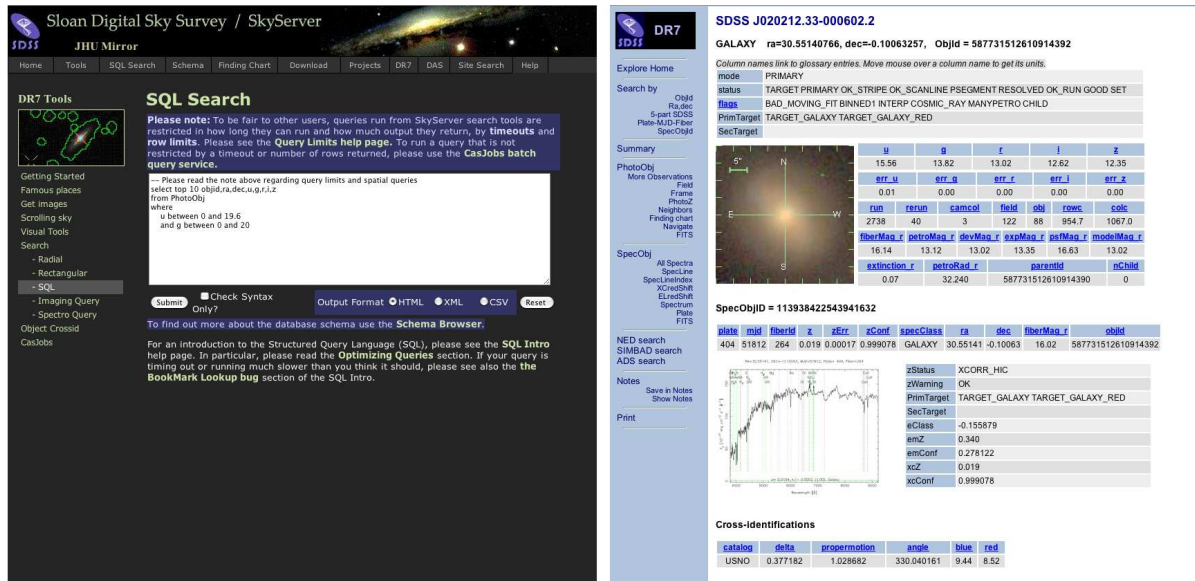


Figura 3.1: Telas do *SkyServer*. À esquerda, formulário para submeter uma *query* SQL. À direita, ferramenta *Explore* mostrando a galáxia NGC 799.

dados do *Data Release 8* do *SDSS* no servidor *CasJobs* do *STARLIGHT*³.

3.1.3 CasJobs

O *Catalog Archive Server Jobs* (*CasJobs*) é um serviço online desenvolvido pela equipe do *SDSS* para expandir a capacidade do *SkyServer* (Li & Thakar 2008). Nele o usuário pode executar consultas SQL no banco de dados do *SDSS* da mesma forma que no *SkyServer*. Porém, além de consultas rápidas, é possível agendar a execução de consultas mais longas. O *CasJobs* gerencia estas consultas agendadas numa fila de execução, de modo a não sobrecarregar a rede ou os servidores de banco de dados. Cada usuário possui sem próprio banco de dados, chamado *MyDB*. Pode-se importar tabelas para o *MyDB* para utilizar em *queries* correlacionando com os dados presentes no *CasJobs*. O *MyDB* serve como armazenamento de tabelas do usuário, e há mecanismos para exportar estas tabelas para arquivos nos formatos FITS, CSV, XML e VOTable. Estes arquivos podem ser lidos por programas de análise de dados como o *TopCat*⁴, ou mesmo importados para outros bancos de dados.

³*CasJobs* do *STARLIGHT*: <http://casjobs.starlight.ufsc.br/casjobs/>

⁴*TopCat* é um visualizador gráfico interativo e editor de dados tabulares usado em astronomia. Ver <http://www.star.bris.ac.uk/~mbt/topcat/>.

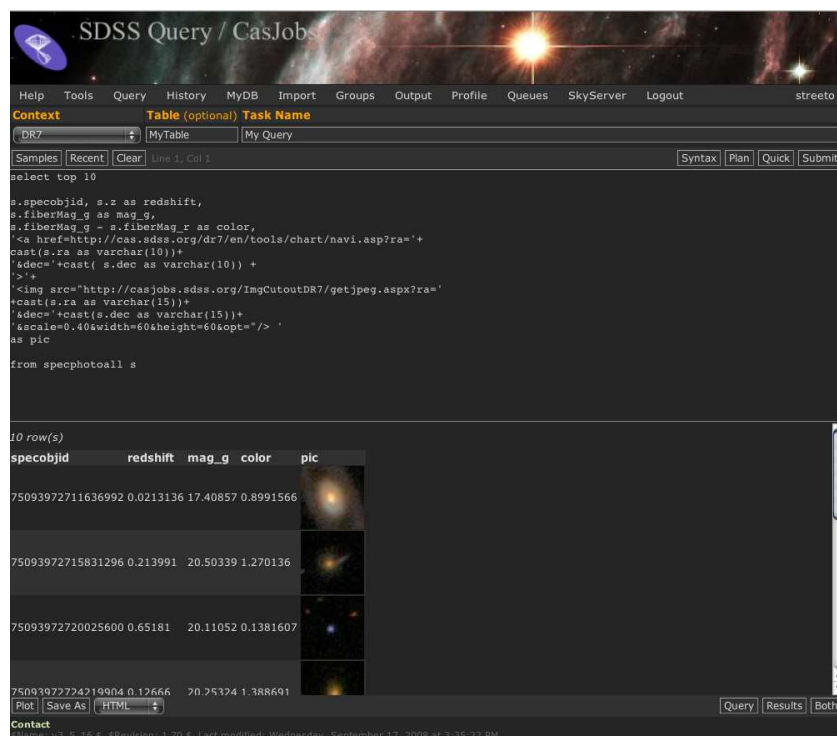


Figura 3.2: Tela do *CasJobs*. Resultado da *query* buscando o *redshift*, a magnitude na banda *g*, a cor *g-r* e uma amostra da imagem de objetos com espectroscopia.

É possível utilizar o *CasJobs* para acessar virtualmente qualquer banco de dados. No momento, o Grupo de Astrofísica da UFSC possui um servidor *CasJobs* com bancos de dados do *STARLIGHT*, *SDSS* DR8, *GalaxyZoo*^[citation needed], e uma amostra do *GALEX* e um catálogo de *redshifts* fotométricos (O’Mill et al. 2011). O *CasJobs* também foi adotado por outros projetos como o *GALEX*, *Kepler*^[citation needed], *Palomar Quest*^[citation needed], *Panoramic Survey Telescope and Rapid Response System* (Pan-STARRS) e até pelo projeto *AmeriFlux*, que contém dados de hidrologia^[citation needed].

A figura 3.2 mostra uma tela típica de uma sessão no *CasJobs*.

3.2 Banco de dados do STARLIGHT

O *STARLIGHT* é um código de síntese espectral (Cid Fernandes et al. 2005). O programa é executado uma vez para cada galáxia do *SDSS*, recebendo seu espectro como um arquivo texto. Ele usa uma biblioteca de espectros de populações estelares simples (*Single Stellar Population*, SSP)⁵ com diferentes idades e metalicidades como uma base do espaço de espectros galácticos possíveis. De forma simplificada, o que o *STARLIGHT* faz é encontrar as frações de massa e luz correspondente a cada elemento da base, ou seja, cada SSP. Analisando as relações entre os componentes determinandos pela síntese, o programa determina diversas propriedades físicas da galáxia, como a massa estelar total, a massa separada por idade, metalicidade média, quantidade de poeira e velocidade de rotação, para citar apenas algumas. Quase um milhão de espectros foram analisados, e o resultado da síntese foi armazenado em arquivos texto.

Apenas os componentes estelares do espectro são obtidos desta forma. Subtraindo o espectro sintetizado de luz estelar é possível medir as linhas de emissão do espectro. Esta é uma etapa de pós-processamento, que gera um catálogo complementar de linhas de emissão.

3.2.1 Importação para o RDBMS

Os arquivos da síntese gerados pelo *STARLIGHT* ocupam uma dezena de gigabytes. Mesmo o catálogo de propriedades físicas das galáxias, sozinho, ocupa mais de um gigabyte. Embora seja um volume razoavelmente grande de dados, é possível trabalhar

⁵Uma SSP consiste num conjunto de estrelas formadas ao mesmo tempo com a mesma metalicidade.^[FIXME]

com esta quantidade de dados num computador atual⁶. A transferência de arquivos com tamanho da ordem de gigabytes pela internet também é lugar-comum atualmente. Poderia-se argumentar que distribuir os dados neste formato seja a forma mais adequada.

Entretanto, deve-se admitir que uma das maiores razões para o sucesso do *CasJobs* em prover acesso aos dados o *SDSS* não é o tamanho da base de dados, e sim a facilidade com que o usuário pode acessar os dados e filtrar apenas o que lhe for conveniente. Além disso, manter a base de dados num local central permite que sejam feitas correções e revisões, o que implicaria normalmente numa nova transferência caso cada usuário tivesse a sua cópia local.

O *CasJobs* requer um servidor rodando *Windows Server* com *Internet Information Services* (IIS) e *Microsoft SQL Server* (MSSQL). A instalação do *CasJobs* está documentada no *website* do *SkyServer* (ver seção 3.1.2). Com um servidor *CasJobs*, o trabalho consiste em importar os dados em arquivos texto para um banco de dados no MSSQL. A ferramenta principal para a manipulação dos bancos de dados no MSSQL é o *Microsoft SQL Server Management Studio*. Nele há um assistente para importação de dados baseado no *SQL Server Integrated Services* (SSIS). A importação das tabelas de propriedades físicas do *STARLIGHT* é trivial. A importação das linhas de emissão requer um trabalho extra para normalizar a tabela⁷.

3.2.2 Estrutura do banco de dados

O esquema do banco de dados do *STARLIGHT* pode ser visto na figura 3.3. A tabela **observational_params** contém os dados obtidos do *SDSS* e usados como parâmetros para a obtenção das propriedades físicas das galáxias. Estas propriedades físicas estão armazenadas na tabela **synthesis_results**. Estas duas tabelas estão relacionadas através da chave *SpecObjID*, que será explicada na seção 3.2.3. A tabela **el_fit** contém medidas de linhas de emissão para cada galáxia, cada linha sendo descrita pelos elementos da tabela **cfg_el_fit**. O manual de acesso aos dados do *STARLIGHT* (anexo A) contém a descrição completa de todos os dados disponíveis nestas tabelas.

⁶Na época da escrita desta dissertação, é difícil encontrar um computador novo com menos de 4 gigabytes de memória RAM.

⁷Normalização de tabelas.^[FIXME]



Figura 3.3: Esquema do banco de dados do *STARLIGHT*.

3.2.3 Amostra do STARLIGHT

A amostra de galáxias do *STARLIGHT* contém 926246 espectros do *SDSS*. A identificação de cada espectro é feita através de um tripleto: a data juliana média da observação (**MJD**, *Mean Julian Date*), a identificação da placa de suporte das fibras ópticas (**Plate**) e a identificação da fibra utilizada para a obtenção do espectro (**FiberID**). Este tripleto (**MJD**, **Plate**, **FiberID**) identifica unicamente um espectro. Porém, é mais conveniente (e eficiente) ter um identificador único⁸ para os registros num banco de dados. No caso do *SDSS*, a tabela de espectros (**SpecObjAll**) tem um identificador chamado **SpecObjID**.

Além de espectros, o banco de dados do *SDSS* (figura 3.4) contém fotometria de 1/4 do céu.^[citation needed] Os objetos com dados de fotometria também tem um identificador único, **ObjID**. Existe uma coluna na tabela de espectros chamada **BestObjID**, que aponta para o registro de fotometria (tabela **PhotoObjAll**) mais provável para cada espectro. É importante salientar que nem todo espectro tem um **BestObjID** definido.

A tabela de índices da amostra de galáxias do *STARLIGHT* (esquema na figura 3.5) contém inicialmente os tripletos [**MJD**, **Plate**, **FiberID**]. Dentro do ambiente *CasJobs* do *SDSS* DR7⁹ a tabela tem os valores de **SpecObjID** e **BestObjID** preenchida através da execução da *query* mostrada na figura 3.6. Entre os objetos na amostra do *STARLIGHT*,

⁸Chave primária [FIXME]

⁹*CasJobs* *SDSS* DR7 - <http://casjobs.sdss.org/CasJobs/>



Figura 3.4: Esquema do banco de dados do *SDSS*.



Figura 3.5: Esquema da tabela de índices da amostra do *STARLIGHT*. Os tipos de dados são referentes à implementação do banco de dados.

622 objetos não tem a sua contrapartida fotométrica.

```

UPDATE sample
    SET SpecObjID=so.SpecObjID, ObjID=so.BestObjID
FROM sample s2 INNER JOIN DR7..SpecObjAll so
    ON so.MJD=s2.MJD
    AND so.Plate=s2.Plate
    AND so.FiberID=s2.FiberID

```

Figura 3.6: Atualização dos índices da amostra de galáxias do *STARLIGHT*. A *query* foi executada no *CasJobs* do *SDSS* DR7 para obter **SpecObjID** e **BestObjID** dado o triploto [MJD, Plate, FiberID].

3.3 Crossmatch SDSS/GALEX

A identificação mútua (*crossmatch*) de objetos em *surveys* diferentes é um problema razoavelmente complicado. A cobertura do céu de cada *survey* em geral não é a mesma. Por outro lado, os objetos presentes em um *survey* podem não ter sido detectados no outro. A probabilidade de duas fontes em catálogos diferentes corresponderem a um mesmo objeto pode ser calculada como função da separação entre elas e a precisão astrométrica das medidas (Budavári & Szalay 2008).

3.3.1 Identificação dos objetos

Budavári et al. (2009) aplicam este método probabilístico ao *SDSS* e ao *GALEX*. O *crossmatch* espacial é feito dentro de um RDMS (MSSQL, o mesmo usado no *CasJobs*), utilizando técnicas avançadas de indexação (seção 3.3.2). A tabela resultante é uma relação “muitos para muitos”, onde a maioria dos objetos *GALEX* tem apenas um objeto *SDSS* associado, mas outras combinações podem ocorrer¹⁰. O artigo apresenta os resultados para dados do *SDSS* DR6 e *GALEX* GR3. Na tabela 3.1 há a quantidade de objetos encontrados nos *surveys* AIS e MIS (estrelas e galáxias) que estão unicamente identificados com um objeto do *SDSS*, utilizando os dados mais recentes (*crossmatch* entre *SDSS* DR7 e *GALEX* GR6, disponível no *CasJobs* do *GALEX*¹¹). A técnica uti-

¹⁰Um exemplo onde pode ocorrer uma combinação “um para muitos” é o caso onde existe uma fonte fraca em UV (presente no *SDSS* mas não detectada pelo *GALEX*) próxima a uma fonte presente tanto no UV quanto no óptico. O algoritmo irá apontar estes dois objetos no *SDSS* como candidatos a serem a contrapartida óptica do objeto detectado no *GALEX*. O caso inverso implicaria numa combinação “muitos para um”

¹¹*CasJobs* do *GALEX*: <http://galex.stsci.edu/casjobs/>

lizada por Budavári et al. agora faz parte do *pipeline* do *GALEX*. A distribuição do *CasJobs* inclui as ferramentas necessárias para fazer o *crossmatch* espacial entre bancos de dados.

Tabela 3.1: Número de objetos do *SDSS* DR7 identificados no catálogo do *GALEX*, por survey e tipo. Apenas são considerados objetos que têm relação “um para um” entre *SDSS* e *GALEX*.

Tipo	AIS	MIS
Galáxia	-	-
Estrela	-	-

3.3.2 Indexação por Hierarchical Triangular Mesh

Para encontrar observações correspondentes de um mesmo objeto em dois *surveys* é necessária uma busca no espaço de coordenadas celestes. Dada a imensa quantidade de objetos presentes tipicamente nos surveys atuais, a verificação de todos contra todos torna-se inviável. (?) aplicaram uma técnica de indexação espacial chamada *Hierarchical Triangular Mesh* (HTM) aos objetos do *SDSS*.

Esta técnica consiste em dividir a esfera celeste numa malha de regiões triangulares (?). Cada uma destas regiões triangulares do céu recebe um identificador único. O processo de divisão em triângulos é recursivo, com a profundidade dependendo da resolução desejada (figura 3.7). O ponto inicial da HTM é um octaedro dado pela intersecção dos eixos x , y e z com a esfera unitária. Os oito primeiros nodos da HTM são os triângulos esféricos compostos por aqueles pontos, conectados por segmentos de grandes círculos (figura 3.8).



Figura 3.7: Todo



Figura 3.8: Todo

3.4 Definição das amostras SDSS/STARLIGHT e GALEX



Figura 3.9: Esquema da tabela de *crossmatch* entre objetos do *GALEX* e do *SDSS*.


```
SELECT INTO mydb..galex_ais
    s.objid AS sdssobjid, x.objid AS galexobjid,
    s.mjd, s.plate, s.fiberid,
    g.nuv_mag, nuv_magErr,
    g.fuv_mag, g.fuv_magErr,
    g.e_bv,
    g.band,
    x.distance,
    pe.nexptime,
    pe.fexptime
FROM mydb..sample s
LEFT JOIN xSDSSDR7 x
    ON s.objid = x.sdssobjid
    AND x.distanceRank=1
    AND x.reverseDistanceRank=1
    AND x.multipleMatchCount=1
    AND x.reverseMultipleMatchCount=1
LEFT JOIN photoobjall g
    ON g.objid = x.objid
LEFT JOIN photoextract e
    ON e.photoextractid=g.photoextractid
WHERE e.mpstype='AIS'
```

Figura 3.10: *Query* para o *match* entre os objetos da amostra do *STARLIGHT* e *GALEX* AIS. A mesma *query* foi usada para o MIS, trocando apenas o nome da tabela para *galex_mis* e modificando a última linha para *e.mpstype='MIS'*.

3.5 Correções na fotometria UV

3.5.1 Poeira

3.5.2 K-correct

Capítulo 4

Problemas astrofísicos

Gaivota com cores UV.

Onde caem as diferentes classes (star forming, retired, AGN, etc, ver ultimo paper do Cid Fernandes et al. (2011)) num diagrama parecido com o de Chilingarian & Zolotukhin (2011).

Capítulo 5

Conclusões e Perspectivas

- 5.1 Este trabalho**
- 5.2 Trabalhos Futuros**

Apêndice A

Anexo 1: Manual de Acesso aos dados do STARLIGHT + Galex

Referências Bibliográficas

- Boksenberg, A., Evans, R. G., Fowler, R. G., Gardner, I. S. K., Houziaux, L., Humphries, C. M., Jamar, C., Macau, D. et al. 1973, MNRAS, 163, 291
- Bowyer, S., Sasseen, T. P., Lampton, M., & Wu, X. 1993, ApJ, 415, 875
- Budavári, T., Heinis, S., Szalay, A. S., Nieto-Santisteban, M., Gupchup, J., Shiao, B., Smith, M., Chang, R. et al. 2009, ApJ, 694, 1281
- Budavári, T. & Szalay, A. S. 2008, ApJ, 679, 301
- Carruthers, G. R. 1973, Appl. Opt., 12, 2501
- Chamberlin, D. D. & Boyce, R. F. 1974, in Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control, SIGFIDET '74 (New York, NY, USA: ACM), 249–264
- Chilingarian, I. & Zolotukhin, I. 2011, ArXiv e-prints
- Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, MNRAS, 358, 363
- Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, MNRAS, 413, 1687
- Codd, E. F. 1970, Commun. ACM, 13, 377
- Code, A. D., Houck, T. E., McNall, J. F., Bless, R. C., & Lillie, C. F. 1970, ApJ, 161, 377
- Henize, K. G., Wray, J. D., Parsons, S. B., Benedict, G. F., Bruhweiler, F. C., Rybski, P. M., & Ocallaghan, F. G. 1975, ApJ, 199, L119
- Kondo, Y. & Wamsteker, W. 1987, Exploring the universe with the IUE satellite, Astrophysics and space science library (D. Reidel)
- Li, N. & Thakar, A. 2008, Computing in Science Engineering, 10, 18
- Maier, D., Stein, J., Otis, A., & Purdy, A. 1986, SIGPLAN Not., 21, 472

- Martin, D. C., Fanson, J., Schiminovich, D., Morrissey, P., Friedman, P. G., Barlow, T. A., Conrow, T., Grange, R. et al. 2005, *ApJ*, 619, L1
- Martin, D. C., Wyder, T. K., Schiminovich, D., Barlow, T. A., Forster, K., Friedman, P. G., Morrissey, P., Neff, S. G. et al. 2007, *ApJS*, 173, 342
- Morrissey, P., Conrow, T., Barlow, T. A., Small, T., Seibert, M., Wyder, T. K., Budavári, T., Arnouts, S. et al. 2007, *ApJS*, 173, 682
- Morrissey, P., Schiminovich, D., Barlow, T. A., Martin, D. C., Blakkolb, B., Conrow, T., Cooke, B., Erickson, K. et al. 2005, *ApJ*, 619, L7
- O’Mill, A. L., Duplancic, F., García Lambas, D., & Sodr , Jr., L. 2011, *MNRAS*, 413, 1395
- Schiminovich, D., Wyder, T. K., Martin, D. C., Johnson, B. D., Salim, S., Seibert, M., Treyer, M. A., Budav ri, T. et al. 2007, *ApJS*, 173, 315
- Stecher, T. P., Cornett, R. H., Greason, M. R., Landsman, W. B., Hill, J. K., Hill, R. S., Bohlin, R. C., Chen, P. C. et al. 1997, *PASP*, 109, 584
- Szalay, A. S., Gray, J., Thakar, A. R., Kunszt, P. Z., Malik, T., Raddick, J., Stoughton, C., & vandenBerg, J. 2002, in *Proceedings of the 2002 ACM SIGMOD international conference on Management of data, SIGMOD ’02* (New York, NY, USA: ACM), 570–581
- Thakar, A. R., Szalay, A. S., Kunszt, P. Z., & Gray, J. 2004, eprint [arXiv:cs/0403020](https://arxiv.org/abs/cs/0403020)
- van Duinen, R. J., Aalders, J. W. G., Wesselius, P. R., Wildeman, K. J., Wu, C. C., Luinge, W., & Snel, D. 1975, *A&A*, 39, 159
- Wyder, T. K., Martin, D. C., Schiminovich, D., Seibert, M., Budav ri, T., Treyer, M. A., Barlow, T. A., Forster, K. et al. 2007, *ApJS*, 173, 293