

JTIMESAT Database: um repositório de dados para séries temporais de imagens de satélite

Fernando Luiz Prochnow Ramme¹
Rubens Augusto Camargo Lamparelli²
Jansle Vieira Rocha¹

¹ Faculdade de Engenharia Agrícola – FEAGRI / UNICAMP
Cidade Universitária “Zeferino Vaz” - Distrito de Barão Geraldo - 13083-875 - Campinas - SP, Brasil
framme@hotmail.com, jansle.rocha@feagri.unicamp.br

² Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura – CEPAGRI / UNICAMP
Cidade Universitária “Zeferino Vaz” - Distrito de Barão Geraldo - 13083-970 - Campinas - SP, Brasil
rubens@cpa.unicamp.br

Abstract. This paper describes the development of a relational database about time series MODIS NDVI 8-day in sugarcane fields in São Paulo State. The data obtained from sugarcane farms consisted of historical data about yields, soil and variety at field level, cartesian location of the each pixel for each subregion monitored, pixel value from red surface reflectance, nir surface reflectance and quality control bands from MODIS images, geographical boundary box for each subregion. All data were integrated in a database. The database was implemented using *Structured Query Language* e *PostgreSQL* relational database management system.

Keywords: remote sensing, image processing, sugarcane, sensoriamento remoto, processamento de imagens, cana-de-açúcar.

1 Introdução

Um significativo conjunto de fenômenos espaciais, como planejamento urbano e a evolução temporal de uma determinada cultura, são inerentemente dinâmicos (dependência da variável tempo). O desenvolvimento de técnicas e abstrações em bancos de dados geográficos que sejam capazes de representar adequadamente fenômenos espaço-temporais dinâmicos vem sendo estudados por diversos autores (Casanova et al., 2005; Câmara et al., 2004; Câmara e Monteiro, 2003; Câmara et al., 1996).

A incorporação da dimensão temporal em um sistema de informação (Câmara, Davis, Monteiro, 2004) não se restringe apenas à questão da representação do tempo, mas inclui também questões relativas a sua recuperação. Um *Sistema de Informações Geográfico (SIG)* é uma excelente ferramenta para investigação de fenômenos diversos relacionados à vegetação (Worboys, Duckham, 2004; Burrough & Mcdonnell, 1998).

A organização de banco de dados geográficos mais utilizada pelas usinas de cana-de-açúcar é, ainda, o modelo geo-relacional. O modelo geo-relacional (Casanova et al., 2005) é constituído de tabelas e arquivos gráficos (vetoriais). As tabelas são, geralmente, armazenadas no formato de arquivo *DBASE - DBF*, por exemplo, para armazenar os atributos dos objetos geográficos. Os arquivos gráficos (como por exemplo, o formato vetorial *Shapefile - SHP*) armazenam as representações geométricas destes objetos. Do ponto de vista do usuário, esta organização permite que aplicações convencionais, concebidas e desenvolvidas dentro do ambiente do *Sistema Gerenciador de Banco de Dados (SGBD)* (Silberschatz, Korth, Sudarshan, 2005), compartilhem os atributos dos objetos geográficos. Os principais problemas desta arquitetura, onde os atributos ficam armazenados no *SGBD* e os vetores estão em arquivos externos ao *SGBD*, são: 1) dificuldade no controle e manipulação das componentes espaciais; 2) dificuldade em manter a integridade entre a componente espacial e a componente alfanumérica; e 3) dificuldade de interoperabilidade, já que cada sistema trabalha com arquivos com formato proprietário.

A fim de superar as dificuldades comumente encontradas no modelo geo-relacional, principalmente referentes a integridade dos dados, este artigo apresenta o resultado do desenvolvimento de um banco de dados relacional, implementado com o objetivo de extrair formas de curvas através do perfil temporal NDVI na cultura da cana-de-açúcar.

2 Material e Métodos

2.1 Dados de localização das regiões a serem monitoradas

A região monitorada no trabalho compreende algumas fazendas localizadas em torno dos municípios Irapuã e Borborema, no estado de São Paulo. A área de estudo possui coordenadas geográficas dos cantos: inferior direito LAT/LONG ($-23,22^\circ$; $-46,86^\circ$) e superior esquerdo LAT/LONG ($-20,69^\circ$; $-49,62^\circ$), na projeção Geográfica, *datum WGS-84* (Figura 1). A cana-de-açúcar (Dillewijn, 1952) é a cultura utilizada como estudo de caso no trabalho.



Figura 1: Localização da área de estudo do trabalho

A fim de armazenar os dados de localização cartesiana, expressa em X e Y, dos pixels selecionados para o monitoramento temporal, o primeiro esquema do banco de dados implementado foi nomeado de *schRegião*. De posse dos vetores, em formato de arquivo *Shapefile*, representativos dos talhões, adquirido da usina de cana-de-açúcar (Usina São Martinho Açúcar e Álcool S/A), com o software de processamento de imagens *ENVI* analisam-se as micro-regiões candidatas e a localização dos pixels “puros”, área do pixel ocupada inteiramente por cana e compreendido em sua totalidade dentro dos talhões agrupados em uma micro-região com características comuns, desejados para o monitoramento em cada micro-região, que posteriormente são armazenados no esquema *schRegião* do banco de dados.

Uma micro-região, neste trabalho, corresponde a um conjunto de talhões com características comuns segundo determinados critérios. Os critérios adotados para a formação de uma micro-região neste trabalho são: 1) talhões contíguos sob o mesmo ambiente (solo e variedade); 2) talhões contíguos sob a presença da cultura da cana-de-açúcar; e 3) período de corte entre os talhões não superior a 8 dias. O critério conservador, a fim de evitar influências na interpretação do comportamento do alvo em estudo, adotado para a seleção dos pixels, na resolução espacial de 250 x 250 metros, consiste em selecionar apenas o pixel que possua cobertura da área superior a 75 % do alvo desejado internamente ao vetor dos talhões.

2.2 Dados de histórico de produtividade da usina de cana-de-açúcar

Nas regiões monitoradas, os dados das safras de cana-soca em 2004/05 e 2005/06 foram obtidos da usina de cana-de-açúcar para cada talhão: identificação do talhão, variedade, solo, fazenda, data de plantio, data de corte, produtividade e estágio da cana. Os dados adquiridos são validados e armazenados no banco de dados estruturado para este trabalho no esquema *schUsina*. A consistência dos dados é requerida porque os dados foram adquiridos no formato tabular (um único arquivo contendo os atributos para cada safra em formato *DBASE*) e vetorial (talhões de cana-de-açúcar, arquivo em formato *Shapefile*).

Problemas comumente encontrados em dados no formato tabular são: a presença de dados incompletos, dados redundantes, dados inconsistentes (como a presença de zero em campos em que de fato ainda não ocorreu a atribuição de valores e/ou valores fora do domínio aceitável conforme as regras de negócio do sistema) e ausência de restrições de integridade de entidade, referencial e de domínio. As restrições de integridade são regras que definem a validade dos dados. Uma base de dados está num estado de integridade quando contém apenas dados válidos e os dados armazenados estão de acordo com a realidade.

A consistência dos dados foi realizada de forma manual. O conceito de temporalidade sobre os dados importados, a partir de cada arquivo tabular em formato *DBASE*, é incorporado durante a importação dos dados, através do preenchimento de tabelas auxiliares que armazenam a identificação do histórico das datas de corte da cana-de-açúcar em cada safra e para cada talhão.

2.3 Dados das imagens de sensoriamento remoto

O *tile H13V11*, do produto Refletância de Superfície MOD09 (MODIS/TERRA), cobre a região de interesse para a cultura da cana-de-açúcar, durante as safras de 2004/05 e 2005/06. As imagens das bandas do vermelho (RED), infravermelho próximo (NIR) e controle de qualidade (QC), versão 4, resolução temporal com composição a cada oito dias, foram adquiridas em USGS/LPDAAC (2007), nos anos de 2004, 2005 e 2006.

Para o processamento, análise dos dados e das imagens MODIS/TERRA foram utilizados os programas: 1) *Modis Reprojection Tool*, para a extração das bandas de interesse, recorte das imagens e re-projeção dos dados para a projeção geográfica LAT/LONG e *datum WGS-1984*, graus decimais; 2) *ENVI*, para a seleção dos pixels desejados na região em estudo; 3) *MATLAB*, para a avaliação de filtros de suavização (Média Móvel, *Savitzky-Golay* e *Wavelets*) sobre a evolução temporal da forma de curva do NDVI; e 4) *PostgreSQL*, como o SGBD.

Os dados de entrada (imagens MODIS) são arquivos (produto MOD09Q1 versão 4) no formato HDF. A versão 4 do produto MOD09Q1 apresenta as seguintes características: tamanho aproximado do arquivo de 275 MB, 4800 colunas x 4800 linhas e Projeção Sinudoidal. Após o recorte das imagens com base na região de estudo do trabalho e na necessidade apenas das bandas de refletância de superfície do vermelho (RED), refletância de superfície do infravermelho próximo (NIR) e controle de qualidade (*Quality Control* ou QC), cada um dos arquivos de imagem de cada banda (composição a cada 8-dias) é armazenado em formato de arquivo *TIFF* (sem compactação, 4,2 MB de tamanho aproximado em disco), contendo 1936 colunas e 1131 linhas, projeção Geográfica LAT/LONG, Datum WGS-84, em graus decimais.

Através de um programa (*script*) desenvolvido neste trabalho, executado sobre o software *MODIS Reprojection Tool*, realiza-se o recorte da região de estudo sobre toda a série temporal de imagens MODIS. Após a finalização desta etapa, a extração dos valores dos pixels das bandas do vermelho (RED), do infravermelho próximo (NIR) e de controle de qualidade (QC) do produto MOD09Q1, nas regiões monitoradas, é realizada de forma automática, via programa nomeado *JTIMESAT TifPixelExtractor* desenvolvido para este trabalho, para toda a

série temporal de dados. Para cada valor de pixel extraído de uma banda de imagem é armazenado juntamente: a identificação da micro-região, o código do produto MODIS, o identificador de posição do pixel e a data da imagem. A extração dos bits da banda QC ocorre posteriormente aos valores dos pixels da banda QC estarem armazenados no banco de dados.

3 Resultados e Discussão

O banco de dados do trabalho *JTIMESAT Database* ou *Java Time Series of Satellite Images Database* é constituído de 3 esquemas: 1) *schUsina*; 2) *schRegião*; e 3) *schPixel*. O modelo do domínio do sistema de informação, em que os principais conceitos que necessitam ser armazenados e persistidos no banco de dados são contemplados, é ilustrado na Figura 2. O objetivo deste modelo é propiciar um fácil acesso aos dados por diferentes SIGs e permitir o armazenamento persistente e a consulta aos dados pelos programas destinados a avaliar a evolução temporal NDVI MODIS.

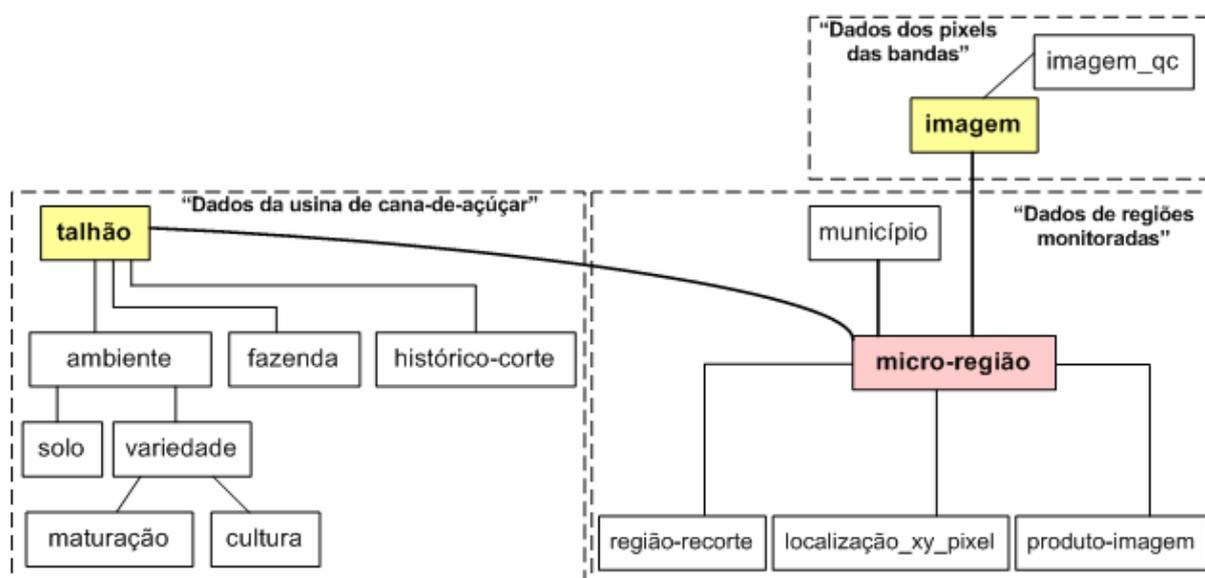


Figura 2: Modelo do domínio do sistema de informação do trabalho *
 * Notação gráfica do diagrama de classes de análise simplificado da *Unified Modeling Language (UML)*

Conforme observado na Figura 2, o esquema *schUsina* é constituído pelos conceitos: talhão, ambiente, fazenda, histórico-corte, solo, variedade, maturação e cultura. A entidade talhão tem por objetivo armazenar o relacionamento entre todas as características de um talhão de cana-de-açúcar temporalmente (através do histórico de corte: data do corte, estágio do corte e produtividade em toneladas de cana por hectare) e contém os dados: identificador único do talhão, identificador único da fazenda associada, identificador único do ambiente de cultivo (caracterizado pelo tipo de solo e variedade da cultura, sendo que a variedade pode estar associada ao tipo de maturação – precoce, média e tardia, e a cultura – cana-de-açúcar), identificador único associado ao vetor do talhão de cana-de-açúcar o qual é característica exclusiva deste vetor em determinada fazenda, código utilizado exclusivamente pela usina de cana-de-açúcar para identificar o talhão, código utilizado exclusivamente pela usina de cana-de-açúcar para identificar a zona do talhão, a data de plantio, o espaçamento (em metros) utilizado no plantio da cultura, a área ocupada pelo talhão em hectares e a área efetivamente plantada em hectares.

Conforme observado na Figura 2, o esquema *schRegião* é constituído pelos conceitos: micro-região, município, região-recorte, localização_xy_pixel e produto-imagem. O conceito de região de estudo neste trabalho compreende a área de recorte das imagens NDVI,

localizada no Estado de São Paulo, de acordo com a Figura 1. A micro-região é uma entidade do esquema *schRegião* que tem por objetivo relacionar um conjunto de talhões com características semelhantes. Uma entidade micro-região contém: um identificador único da micro-região, um identificador único do município, as coordenadas geográficas do pixel central do conjunto de talhões contidos na micro-região, as coordenadas geográficas do canto superior direito e do canto inferior esquerdo de um retângulo que contorna a micro-região. Uma micro-região é, portanto, constituída de um ou mais talhões com características comuns. Cada talhão armazenado neste banco de dados possui um ou mais pixels com coordenadas escalares (localização X e Y na imagem da região de recorte). A entidade produto-imagem consiste em armazenar a descrição das características das seis bandas das imagens MODIS/TERRA utilizadas no trabalho: "MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid, RED Band" (MOD09Q1.004.0250m.RED), "MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid, NIR Band" (MOD09Q1.004.0250m.NIR), "MODIS/Terra Quality Control 8-Day L3 Global 250m SIN Grid, QC Band" (MOD09Q1.004.0250m.QC), "MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid, RED Band" (MOD09Q1.005.0250m.RED), "MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid, NIR Band" (MOD09Q1.005.0250m.NIR), "MODIS/Terra Quality Control 8-Day L3 Global 250m SIN Grid, QC Band" (MOD09Q1.005.0250m.QC). No esquema *schRegião* também é possível armazenar a definição de outros sensores, satélites e produtos.

Através de uma simples consulta aos dados contidos no banco de dados, através da linguagem *Structured Query Language (SQL-92)*, é possível obter quantos pixels estão associados a uma determinada micro-região (*oid_microregiao*) e a uma determinada banda da imagem (*cod_produtosig*), conforme comando a seguir.

```
SELECT count(*)
FROM bdregiao.tpixellocal
WHERE oid_microregiao=1 AND
cod_produtosig='MOD09Q1.004.0250m.RED' ;
```

Outra combinação de comando de consulta *SQL-92* ao esquema *schRegião* a fim de obter quais são os talhões de cana-de-açúcar relacionados a uma determinada entidade micro-região é apresentado abaixo.

```
SELECT oid_talhao
FROM bdregiao.tmicroregiao_ttalhao
WHERE oid_microregiao=1;
```

Conforme observado na Figura 2, o esquema *schPixel* é constituído pelos conceitos: imagem e imagem_qc. A entidade imagem é responsável por armazenar o valor de cada pixel para cada banda da imagem MODIS, e relacioná-lo: ao identificador único da micro-região, ao código do produto-imagem, a um identificador incremental de seqüência do pixel na micro-região e a data da imagem MODIS. A entidade imagem_qc é responsável por armazenar os valores de controle de qualidade (*Quality Control - QC*) avaliados posteriormente de cada pixel da banda QC das imagens MODIS. As seguintes características de qualidade do pixel são armazenadas na entidade imagem_qc para posterior utilização em filtragem dos ruídos: avaliação da qualidade de imageamento do pixel da superfície da terra, qualidade referente a presença de nuvens, qualidade de imageamento do pixel na banda vermelho, qualidade de imageamento do pixel na banda infravermelho próximo, qualidade da correção atmosférica e a qualidade da correção dos efeitos provocados pelos pixels da adjacência.

A combinação de comando de consulta *SQL-92* ao esquema *schPixel* a fim de obter o valor do NDVI (valor médio dos pixels contidos em uma determinada entidade micro-região

oid_microregiao sem o critério de filtragem de dados) para uma determinada data (*dataimg*) é apresentado abaixo.

```

SELECT AVG(
  ((CAST((pixels_nir.pixel_valor - pixels_red.pixel_valor) AS FLOAT8))/
   (CAST((pixels_nir.pixel_valor + pixels_red.pixel_valor) AS FLOAT8))
  ) ) as ndvi
FROM
  (SELECT oid_pixelpos, pixel_valor
   FROM bdpixel.timagem
   WHERE oid_microregiao=2
        AND cod_produtosig='MOD09GQK.004.0250m.RED'
        AND dataimg='2005-05-21' ORDER BY oid_pixelpos
  ) AS pixels_red,
  (SELECT oid_pixelpos, pixel_valor
   FROM bdpixel.timagem
   WHERE oid_microregiao=2
        AND cod_produtosig='MOD09GQK.004.0250m.NIR'
        AND dataimg='2005-05-21' ORDER BY oid_pixelpos
  ) AS pixels_nir
WHERE pixels_red.oid_pixelpos = pixels_nir.oid_pixelpos;

```

O banco de dados desenvolvido para o trabalho contempla o relacionamento entre 1) os dados obtidos da usina de cana-de-açúcar (dados de plantio e colheita nos talhões de cana-de-açúcar); 2) os dados de localização cartesiana dos pixels selecionados para cada micro-região; e 3) dados dos valores extraídos das bandas RED, NIR e QC do produto MODIS/TERRA MOD09Q1 versão 4. Conforme observado na Figura 2, os dados adquiridos das usinas de cana-de-açúcar são armazenados no esquema *schUsina*. Uma usina de cana-de-açúcar pode possuir várias fazendas sob monitoramento e cada uma dessas fazendas está relacionada a um conjunto de talhões. Em cada agrupamento de talhões (talhões contíguos) tem-se uma micro-região de estudo. O conjunto de micro-regiões é armazenado no esquema *schRegião*. No esquema *schRegião*, a localização de cada pixel de cada micro-região é armazenada para fins de extração dos dados de nível de cinza (NC) de cada banda da imagem utilizada no trabalho. O esquema *schPixel* armazena os dados de NC de cada pixel das bandas RED, NIR e QC, temporalmente. A extração dos dados de NC de cada pixel das bandas RED, NIR e QC, temporalmente, é realizada de forma automática através de software desenvolvido neste trabalho denominado *JTIMESAT TifPixelExtractor*.

A fim de obter uma infraestrutura de acesso aos dados desvinculada a qualquer SIG e, conseqüentemente, de uma linguagem de programação proprietária do SIG, desenvolveu-se neste trabalho o projeto e a implementação do banco de dados *JTIMESAT Database*. A Figura 3 apresenta o diagrama de interoperabilidade *JTIMESAT Database* com os diversos softwares, incluindo o SIG. Através de um SIG, usualmente é possível comunicar-se com o SGBD através da linguagem de consulta padrão SQL-92. Entre as vantagens em adotar um banco de dados ao invés de tabelas com atributos, a fim de realizar o desacoplamento dos dados com um software SIG, destacam-se: 1) a independência da linguagem de programação para implementar algoritmos de maior complexidade a fim de manipular os atributos; 2) acesso concorrente aos dados por múltiplos softwares; 3) acesso persistente aos dados; 4) menor volume de dados, onde apenas os valores dos pixels selecionados são armazenados no banco de dados, ao invés de uma série de imagens das bandas RED, NIR e QC; 5) maior flexibilidade em utilizar softwares diversos quanto ao acesso, manipulação, inspeção e geração de relatórios a partir dos dados contidos no banco de dados; 6) menor necessidade de recursos computacionais de processamento; e 7) maior agilidade no processamento dos dados e extração automática de formas de curvas.

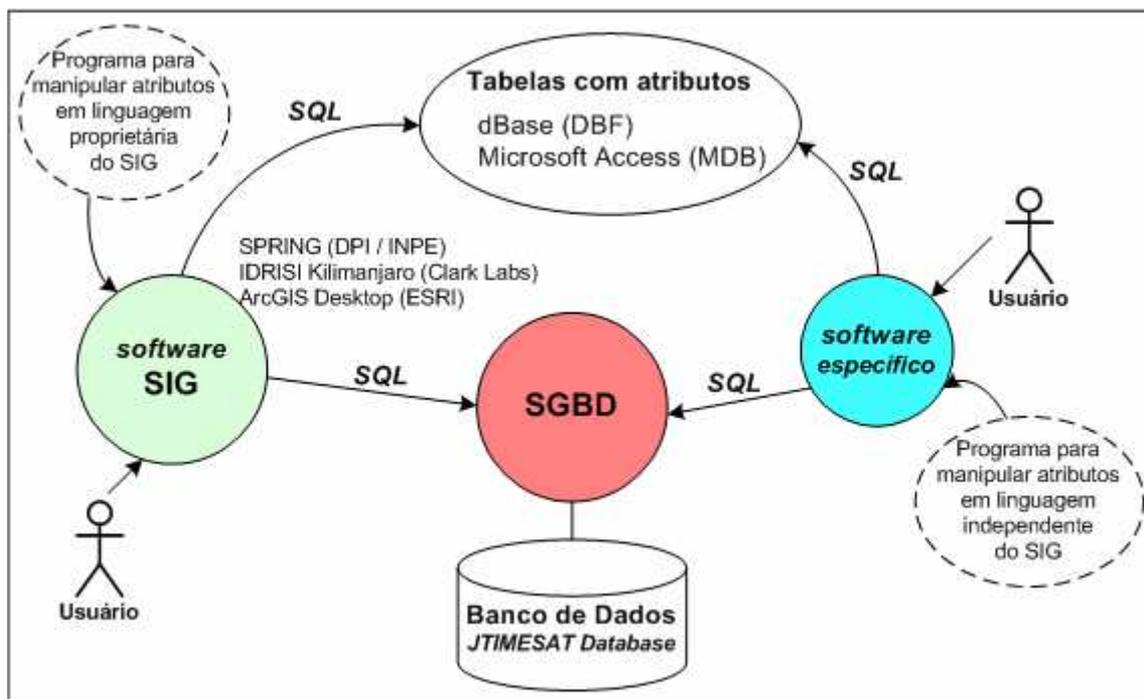


Figura 3: Diagrama de interoperabilidade *JTIMESAT Database* com os software diversos

4 Conclusões

O banco de dados *JTIMESAT Database* implementado possibilita extrair formas de curvas através do perfil temporal NDVI na cultura da cana-de-açúcar. A modelagem concebida para o banco de dados é de fácil entendimento para o usuário, propicia integridade aos dados, possibilita obter alto desempenho para as consultas aos dados, reduz o espaço necessário em disco para o armazenamento dos dados, atende aos requisitos do estudo de caso deste trabalho e é acessível por qualquer software, incluindo os SIGs, através da linguagem de consulta padronizada *SQL-92*.

Referências Bibliográficas

- BURROUGH, Peter A.; MCDONNELL, Rachael A. **Principles of geographical information systems**. New York: Oxford University Press Inc., 1998.
- CASANOVA, Marco Antônio; et al.. **Bancos de dados geográficos**. Curitiba: MundoGEO, 2005.
- CÂMARA, Gilberto; et al. **Análise espacial de dados geográficos**. Brasília: EMBRAPA, 2004.
- CÂMARA, Gilberto; DAVIS, Clodoveu; MONTEIRO, Antônio Miguel Vieira. **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2004.
- CÂMARA, Gilberto; MONTEIRO, Antônio Miguel Vieira. **Introdução à modelagem dinâmica espacial**. São José dos Campos: INPE, 2003.
- CÂMARA, Gilberto; CASANOVA, Marco Antônio; MEDEIROS, Claudia Maria Bauzer; et al. **Anatomia de sistemas de informação geográfica**. Campinas: Instituto de Computação, Unicamp, 1996.
- DILLEWIJN, Cornelis Van. **Botany of sugarcane**. Waltham: Chronica Botânica Co., 1952.
- GSFC/NASA. **Moderate Resolution Imaging Spectroradiometer (MODIS): technical specifications**. Disponível em: <<http://modis.gsfc.nasa.gov/about/design.php>>. Acesso em: 5.nov. 2008.

PostgreSQL. **PostgreSQL 8.2 documentation**. Disponível em: <<http://www.postgresql.org/>>. Acesso em: 18.jan.2008.

SILBERSCHATZ, Avi; KORTH, Henry F.; SUDARSHAN, S. **Database System Concepts**. 5.ed. New York: McGraw-Hill, 2005.

Usina São Martinho Açúcar e Álcool S/A. Disponível em: < www.usinasaomartinho.ind.br >. Acesso em: 5.nov. 2008.

USGS/LPDAAC. **Imagens dos satélites MODIS/TERRA e MODIS/AQUA, disponibilizadas por U.S. Geological Survey / Land Processes Distributed Active Archive Center**. Disponível em: < <http://lpdaac.usgs.gov/modis/dataproducts.asp>>. Acesso em: 16.abr.2007.

WORBOYS, Michael F.; DUCKHAM, Matt. **GIS: a computing perspective**. 2 ed. New York: CRC Press, 2004.