INPE ePrint: sid.inpe.br/mtc-m19@80/2010/08.03.13.46 v1 2010-08-04

Sistema de Processamento de Dados de Radiação Solar (SPDRAD)

Leandro de Oliveira Macedo [*,**] e Juan Carlos Ceballos [*]

[*] Divisão de Satélites e Sistemas Ambientais, CPTEC, INPE – Cachoeira Paulista – SP.
[**] FATEC, Guaratinguetá – SP.
e-mails: leandro.macedo@cptec.inpe.br, juan.ceballos@cptec.inpe.br

ABSTRACT. SPDRAD (Solar Radiation Data Processing System) is an assembly of several systems: (a) MatLab programs as front end, which communicates with (b) a SGBD (Data Base Management System) in PostGRESql, which contains general metadata and tabulated data (radiation data from several ground networks, as well as thevalues estimated by model GL/CPTEC based on satellite imagery), also allowing to extract information from (c) an extended data collection of binary files (mainly GL high resolution products), from a Terabyte Storage Disk. The system is in continuous development. Its structure and some preliminary tasks are shown.

1 - INTRODUÇÃO

Atualmente a DSA é a principal responsável pelas operações dados gerados por satélites, no âmbito do CPTEC, além de ser responsável pela geração e manutenção de diversos produtos.

O modelo GL (GLobal radiation model) versão 1.2, é um produto DSA que estima radiação solar por satélite (área de cobertura atual se estende desde 100ºW até 25ºW e desde 50ºS a 13ºN, com resolução espacial de 0.04º, caracterizando uma matriz de 1800x1800 pixels) sobre a América do Sul baseado em imagens do sensor VIS a bordo do satélite GOES. Os dados GL necessitam ser comparados com outras fontes de dados, a fim de avaliar a precisão do modelo. Neste caso, são utilizados dados fornecidos por diversas redes de estações de superfície, pertencentes a diversas instituições.

O volume de dados gerados pelo modelo e sua conseqüente validação é muito alto. Devido a isto, os dados mais antigos são armazenados em fitas do tipo DAT e LTO, estando disponível nos servidores de arquivos da divisão apenas a última semana de dados. Por serem mídias de acesso seqüencial, necessita-se transferir o conteúdo de toda a fita (o tamanho varia de 80 a 280 GB, dependendo do tipo de fita) para um disco rígido, para só então resgatar determinado dado. Por outro lado, os arquivos armazenados nos servidores não possuem uma organização padrão, obrigando o usuário a "garimpar" a rede de toda a divisão em busca dos dados que necessita.

Deste modo, identificou-se a necessidade de organizar os dados do modelo GL e de estações em uma estrutura computacional (SPDRAD) que ofereça: 1) centralização dos dados de radiação solar, facilitando sua localização; 2) recuperação ágil destes dados; 3) possibilidade de manipulação e análise dos mesmos. A seguir, apresentamos o esquema de estrutura concebido e alguns resultados preliminares.

2 - A ESTRUTURA COMPUTACIONAL PROPOSTA

Para centralizar os dados GL e oferecer facilidade de aquisição e manipulação dos dados:

- É necessário um espaço de armazenamento destinado a guardar arquivos binários gerados pelo modelo GL dentro de uma estrutura de diretórios padrão;
- É necessário armazenar os dados de estações de superfície disponíveis na DSA de forma a oferecer uma base de dados para comparações com dados estimados pelo modelo GL;
- É necessário oferecer rotinas de recuperação e manipulação desses dados e através de uma interface que permita ao usuário realizar tais tarefas.

Baseado nestes requisitos, a solução proposta é um ambiente computacional composto de: a) um disco rígido com elevada capacidade de armazenamento, destinado a guardar arquivos binários de radiação; b) um sistema do tipo SGBD (Sistema Gerenciador de Bancos de Dados) para armazenar dados medidos por solarímetros, resultados GL nos locais de medição e informações geográficas destes locais; c) e uma aplicação (SIRAD) com interface simples e que forneça as ferramentas de pesquisa, aquisição e análise dos dados e arquivos disponíveis.

2.1. O espaço de armazenamento de arquivos

Consiste em um disco que possui um espaço de 2 TB destinado a armazenar arquivos binários de saída do modelo (arquivos UV+VIS e Global média diária e arquivos de refletância horários). Este disco (*nickname* MAGALI, pela sua capacidade de "engolir" dados) está sendo alimentado com a série histórica (cerca de 15 anos de dados) de binários GL. Estes binários estão organizados em uma estrutura padrão de diretórios. Esta estrutura é semelhante à mostrada na figura 1. Estes binários estão compactados em arquivos ZIP, e utiliza-se o protocolo de transferência de arquivos (FTP) para acesso aos binários, contudo este acesso é permitido apenas através da aplicação de interface.

Rotinas auxiliares de mineração de dados (Data Mining) e também de ILM (Information LifeCycle Management) estão sendo estudadas para aumentar a eficiência das pesquisas de arquivos e o controle inteligente do espaço de armazenamento, respectivamente.

2.2. O Banco de dados de Radiação Solar

Consiste de um sistema do tipo SGBD destinado a armazenar resultados pontuais extraídos de binários GL e medidos por estações de superfície ([1]), além de armazenar informações geográficas dos pontos de medição (id, latitude, longitude, município, nome, dono) e metadados dos arquivos binários (dimensões, resolução, lat/lon inicial e final, localização do arquivo). Também armazena as planilhas de Status, e Estatísticas descritas em Galvão e Ceballos (2008), sob a forma de tabelas, sendo estas, parte da estrutura de controle na validação mensal do modelo. As principais vantagens em se utilizar um SGBD estão na recuperação ágil de grandes volumes de registros, e na possibilidade de "combiná-los" gerando relatórios diversos.

[1] Recebe-se um fluxo contínuo de dados de mais de 350 PCDs (estações meteorológicas automáticas) dos Sistemas Estaduais de Meteorologia (tri-horários), armazenados no CPTEC, além de cerca de 446 PCDs do INMET (dados horários) e dezenas de sítios de outras redes.

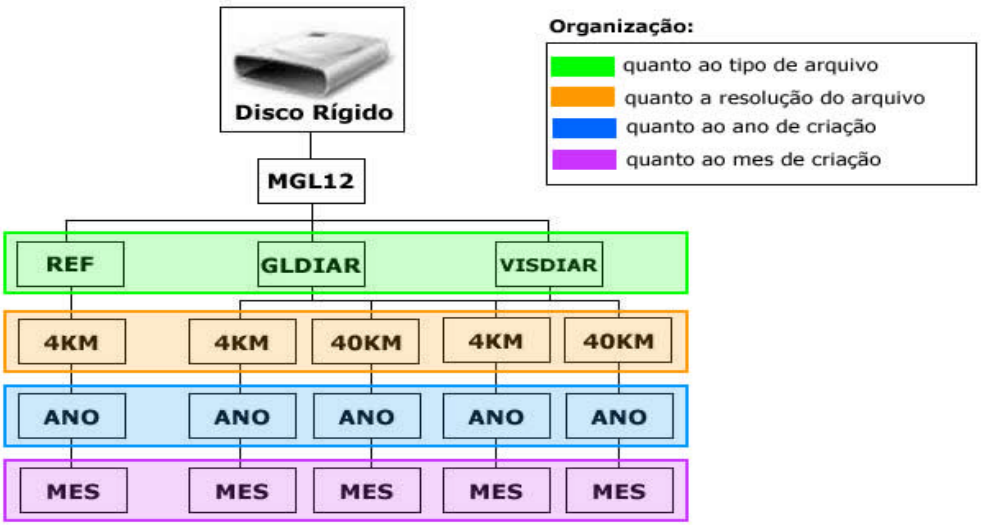


Figura 1 – Estrutura de diretórios do disco MAGALI

2.2. O Banco de dados de Radiação Solar

Consiste de um sistema do tipo SGBD destinado a armazenar resultados pontuais extraídos de binários GL e medidos por estações de superfície ([2]), além de armazenar informações geográficas dos pontos de medição (id, latitude, longitude, município, nome, dono) e metadados dos arquivos binários (dimensões, resolução, lat/lon inicial e final, localização do arquivo). Também armazena as planilhas de Status, e Estatísticas descritas em Galvão e Ceballos (2008), sob a forma de tabelas, sendo estas, parte da estrutura de controle na validação mensal do modelo. As principais vantagens em se utilizar um SGBD estão na recuperação ágil de grandes volumes de registros, e na possibilidade de "combiná-los" gerando relatórios diversos.

O SGBD proposto é o PostgreSql, por oferecer robustez e elevada performance, além de ser o Banco de Dados disponível em ambiente operacional dentro da DSA. Utiliza-se também a extensão PostGIS, que é um conjunto de APIs que possibilitam a criação e utilização de Bancos de dados PostgreSql em modo geográfico. Esta extensão está instalada no banco de dados de testes visando uma futura integração do SPDRAD com o software SIGMA (Sistema de Informação Geográfica para o Meio Ambiente) da DSA. Um esquema do Banco de dados de radiação solar (BDRAD) é mostrado na figura 2.

As entidades presentes nesta fase do BDRAD são:

- **Sites** – Tabela de informações sobre os sites de amostragem
- **Rad Diar** – Tabela que armazena informações pontuais de radiação solar do modelo e estações.
- **Status_Data –** Tabela que armazena flags de análise dos dados GL e das estações.
- **Estatísticas –** Tabelas com resultados estatísticos da comparação GL/estação.
- **Diferenças** – Tabela com os valores de diferença entre modelo e estação.

As demais entidades são tabelas que armazenam informações adicionais (como países que possuem sites, donos das redes, Estados de cada país).

[2] Recebe-se um fluxo contínuo de dados de mais de 350 PCDs (estações meteorológicas automáticas) dos Sistemas Estaduais de Meteorologia (tri-horários), armazenados no CPTEC, além de cerca de 446 PCDs do INMET (dados horários) e dezenas de sítios de outras redes.

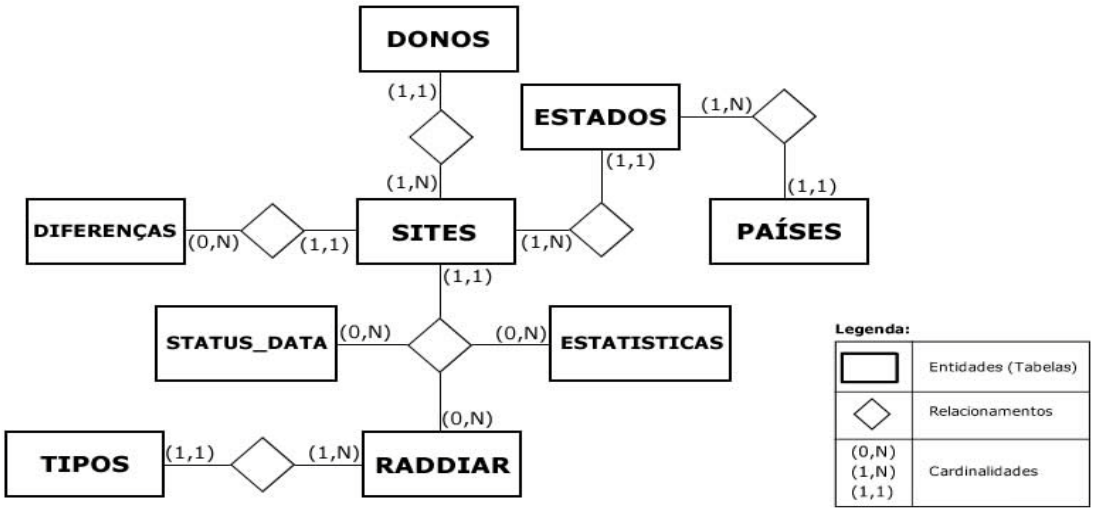


Figura 2 - Esquema conceitual do Banco de Dados de Radiação Solar

2.3 - A Aplicação e Interface da Estrutura Proposta

Consiste de um conjunto de funções e rotinas, e uma interface, escritas em linguagem computacional Matlab, que fornecem os meios necessários para manipulação e análise dos dados armazenados no disco (MAGALI) e no BDRAD. A esta aplicação atribui-se o nome de Sistema de Interface de Recuperação e Avaliação de Dados (SIRAD).

O SIRAD está sendo desenvolvido em Matlab, por este oferecer suporte ao processamento de imagens assim como funções de análise estatística de dados e ferramentas de manipulação de arquivos. Uma descrição detalhada das ferramentas Matlab encontra-se em Hanselman e Littlefield (2003) e na internet (http://www.mathworks.com/products/matlab/, acessado em maio de 2010).

Uma descrição dos critérios de processamento de dados GL encontra-se em Galvão e Ceballos (2008). Macedo e Ceballos (2009) aplicaram estes critérios na comparação de resultados baseados em imagens GOES e Meteosat. No SPDRAD, a interface com o usuário inclui um conjunto de janelas (GUIs) onde o usuário tem a possibilidade de manipular e visualizar de maneira simples os dados e análises processadas pela aplicação.

3 – RESULTADOS PRELIMINARES

Foi implementado um esquema piloto do SIRAD aplicado a outubro de 2009, testando as seguintes funcionalidades: 1) procura de dados diários do modelo e de estações; 2) resgate de séries mensais de locais específicos e de regiões; 3) resgate de dados pontuais dos binários armazenados na MAGALI; 4) análise estatística dos resultados GL comparados com medições de estações.

Para local e data especificados pelo usuário, o SIRAD exibe um breve resumo da estação (ID, latitude, longitude, município, dono) assim como as irradiâncias medidas e estimada (Figura 3), permitindo um primeiro monitoramento do modelo e da rede. Para o mesmo local, a aplicação recupera a série mensal do modelo e da estação num gráfico de barras, mais desvio médio mensal e desvio padrão (Figura 4a).

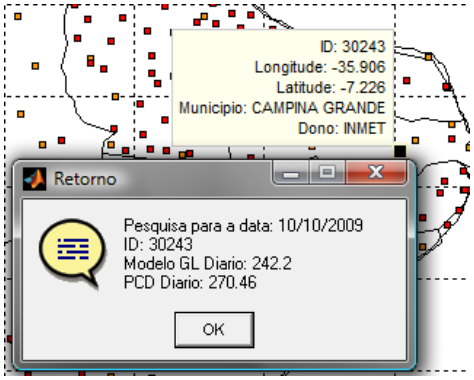
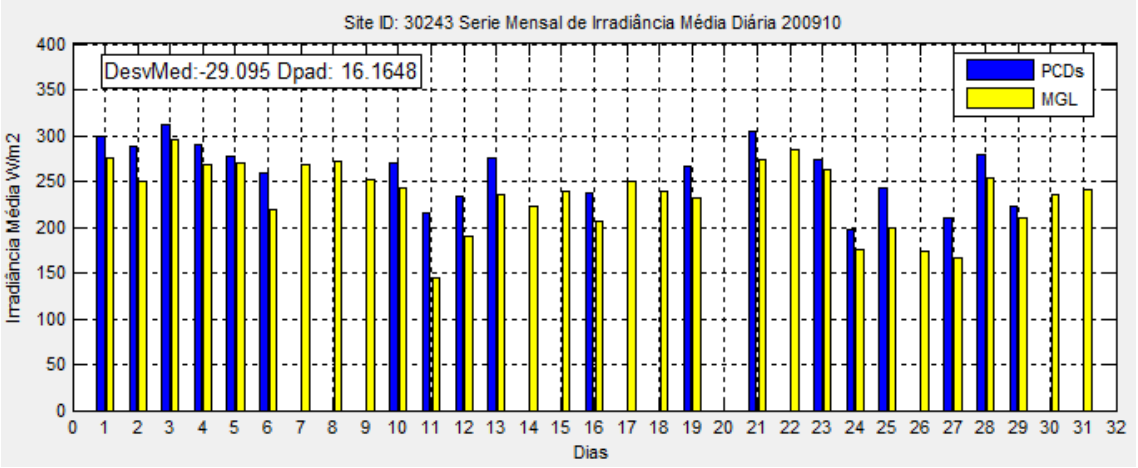
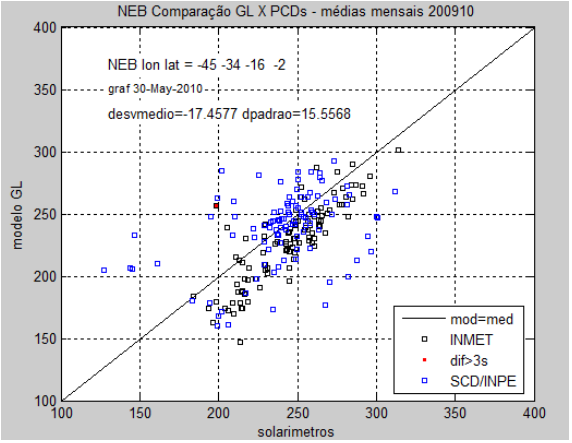


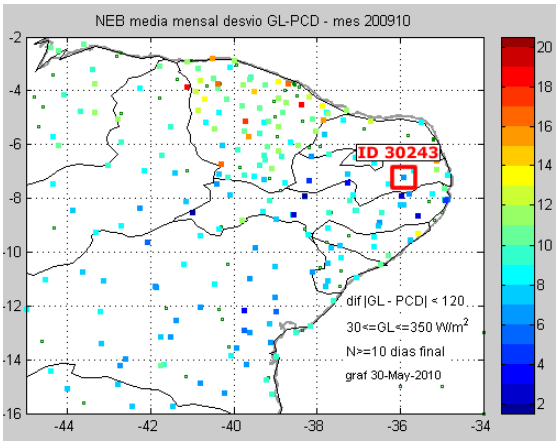
Figura 3 - Exemplo de visualização de informações SIRAD



a) Exemplo de série mensal pesquisada e exibida pelo SIRAD.



b) Dispersão GL x PCD avaliada pelo SIRAD



c) Desvios médios mensais nos locais amostrados na região Nordeste

Figura 4 – Exemplos de análises realizadas pelo SIRAD.

O sistema ainda permite a recuperação de dados pontuais diretamente dos binários, quando não forem encontrados no SGBD. Isto permite uma utilização mais

abrangente dos dados GL, uma vez que o usuário pode resgatar dados de qualquer local dentro das dimensões do arquivo (1800 x 1800, com resolução de 0.04º por pixel).

Uma análise da qualidade do modelo em outubro de 2009 foi obtida utilizando a ferramenta de validação mensal presente na aplicação. Os resultados preliminares indicam que o modelo representa bem a distribuição de radiação solar no Brasil para outubro de 2009. Para a região Nordeste (escolhida também como exemplo piloto), o desvio médio mensal é da ordem de 17 W. m^{-2} e o desvio padrão é de 15 W. m^{-2}. O SIRAD exibe um gráfico de dispersão que permite visualizar o ajuste GL-solarímetros (figura 4b), e também exibe um gráfico do desvio médio mensal na região amostrada (Nordeste do Brasil, ver figura 4c). Esses números indicam um bom ajuste entre dados GL e de estações para essa região.

4 – CONCLUSÕES

O SPDRAD mostra-se de grande utilidade tanto na organização dos dados de radiação solar, quanto na facilidade de acesso e manipulação dos mesmos. Será de grande utilidade para complementar-se com resultados de pesquisa conexa (especialmente o monitoramento de saldos de radiação por satélite), assim como para integrar-se como fonte de dados de radiação no Banco de Dados e Imagens da Divisão de Satélites Ambientais do CPTEC. Ainda, novas funções poderão ser adicionadas sem grandes mudanças estruturais.

Se comprovada a eficiência do sistema proposto, sua base conceitual poderá ser aplicada para outros produtos da Divisão, contribuindo para gerar um ambiente colaborativo de pesquisa e aquisição de dados gerados na DSA.

Agradecimentos. Este trabalho foi desenvolvido no marco do Programa PIBIC INPE/CNPQ, através de bolsa de iniciação científica, com apoio do projeto CNPq-Milênio Processo 484862/07-3

5 – REFERÊNCIAS

Galvão, A. M. e Ceballos, J. C. **Sistema de gerenciamento e validação de dados de radiação solar**. XV Congresso Brasileiro de Meteorologia, **Anais** (CD-ROM), São Paulo, 2008.

Hanselmann, D. e Littlefield, B. **MatLab 6 – Curso Completo**. Prentice Hall, São Paulo, 2003, 676 pp.

Macedo, L. O. e Ceballos, J. C. **Análise da qualidade da estimativa de radiação solar do modelo GL/CPTEC utilizando imagens SEVIRI-MSG.** Relatório de Iniciação Científica PIBIC-INPE, julho 2009. 28 pp.

Manual de Referencia do PostgreSQL 8.0.0. 2008. Acessado em 29 de maio de 2010. Disponível em: http://pgdocptbr.sourceforge.net/pg80/index.html