

ATON- SOFTWARE DE AQUISIÇÃO, GERENCIAMENTO E COMPARTILHAMENTO DE DADOS DE RADIOMETRIA SOLAR.

Miguel Suarez Xavier Penteado – miguel_penteado@fca.unesp.br

Universidade Estadual Paulista, Faculdade de Ciências Agrônomicas, Departamento de Recursos Naturais

João Francisco Escobedo – escobedo@fca.unesp.br

Universidade Estadual Paulista, Faculdade de Ciências Agrônomicas, Departamento de Recursos Naturais

Alexandre Dal Pai – aadalpai@fatecbt.edu.br

Centro Paula Souza, Faculdade de Tecnologia de Botucatu

Resumo. Este documento apresenta o desenvolvimento do Sistema Aton, software livre de aquisição, tratamento e compartilhamento de dados de radiometria solar no formato compatível com ao software de integração de bases de dados científicas OpenDap. Aton também conta com funcionalidades de correção de séries de dados danificadas por causas mecânicas ou elétricas nos sensores e dataloggers.

Palavras-chave: Energia Solar, SQL, OpenDap, netCDF, Aton

1. INTRODUÇÃO

O estudo do comportamento dos fenômenos meteorológicos é de fundamental importância para humanidade, assim como se tem constatado recentemente com a acentuada mudança climática global. O levantamento de evidências científicas é necessário para justificar tomadas de decisões, seja na esfera governamental ou cotidiana. Contudo, sabe-se que as efemérides meteorológicas envolvem um grande número de variáveis e parâmetros, dificultando o seu processamento por ferramentas matemáticas, tais como a estatística e a análise probabilística devido ao seu imenso volume, sem o auxílio da automação. Em consequência, a construção e validação de modelos que representem e antecipem acontecimentos climáticos se torna muito mais lenta, na melhor hipótese. Portanto, o ingresso da informática nas ferramentas de análise científica torna-se praticamente um requisito.

Neste contexto a palavra informática remete a ciência que estuda conceitos fundamentais de computação: a ciência da computação. Um grande esforço neste sentido já foi e ainda está sendo feito. O University Corporation for Atmospheric Research (UCAR), consórcio criado por 14 universidades norte-americanas na década de 50 do século passado, que atualmente compreende 71, desenvolveu segundo BALTUCH (1997), um projeto de integração e disponibilização de dados de sensoriamento climatológico acessível via internet. O trabalho teve início em 1982 dentro de um programa de metas chamado unidata, que propôs e desenvolveu a estrutura de integração de bases de dados chamada Internet Data Distribution (IDD). O IDD possibilita publicar e integrar os bancos de dados climatológicos obtidos em dispositivos de sensoriamento localizados em laboratórios das universidades que participam do consórcio. Entre as bases de dados disponíveis criadas por esta infra-estrutura, pode-se citar a de sensoriamento de terremotos, a de campo eletromagnético terrestre e a de efemérides oceanográficas. Contudo o programa IDD está vinculado ao consórcio UCAR. Estas bases de dados podem ser concatenadas e integradas graças ao padrão de formato de dados científicos inicialmente proposto e desenvolvido pela NASA denominado Common Data Format (CDF), e mais recentemente seu sucessor, o netCDF (REW e DAVIS, 1990). netCDF é um formato binário de dados portátil, que suporta acesso direto, compartilhamento, pós-inserção e metadados.

Softwares desenvolvidos pelo programa de metas unidata e ligados a infra-estrutura IDD extraem e agrupam os dados dos sensores neste formato. O uso do formato netCDF possibilita as bases de dados serem estocadas, movidas e manipuladas local ou remotamente através do auxílio dos softwares desenvolvidos pelo unidata. Contudo, apensar da estrutura estar disponível a mais de uma década, a mesma não é aproveitada para estocagem e compartilhamento de dados de sensoriamento de radiometria solar gerado nos laboratórios em território brasileiro. Cabe ressaltar que este formato sofre de algumas limitações se comparado ao modelo estocagem em banco de dados relacional (SILBERSCHATZ e KORTH, 1999:1). Não existe o conceito de restrições de integridade imposta aos dados (SILBERSCHATZ e KORTH, 1999:28), não há suporte a transações (DATE, 2000:77) e por consequência deve haver um esforço extra de programação para utilizá-lo em aplicações que requerem gravação simultânea de usuários distintos na mesma base de dados.

O formato netCDF utiliza matrizes de dados multidimensionais como forma de armazenamento de dados no lugar de tabelas, o que dificulta, (mas não impossibilita), a estocagem de bases de dados neste formato em Sistemas de Banco de Dados Relacionais, devido ao fato de tabelas de Bancos de Dados serem na verdade matrizes de dados de dimensão

2. Isso dificulta o emprego de técnicas de Mineiração de Dados, um conceito que possibilitaria a busca por padrões de comportamentos de causas solares comuns através da análise de efeitos em bases de dados de áreas de concentração distintas da climatologia.

O controle da aquisição de dados, precisão dos mesmos e a busca por padrões de comportamento dentro de séries de medidas são tarefas rotineiras em laboratórios de análise. Segundo INMON (2001:16), "... reunião de padrões é o começo de uma interessante seqüência de eventos. Uma vez que os dados reunidos e que os padrões foram analisados, esses padrões são identificados e usados como base para um comportamento compreensível. Uma vez que o comportamento é entendido, ele pode ser previsto. Uma vez que o comportamento pode ser previsto, o posicionamento da corporação [dos interessados, dos pesquisadores] é possível." Sempre um posicionamento busca uma situação favorável que leva a uma oportunidade. Portanto pode-se aproveitar oportunidades uma vez que estas são reconhecidas. Analise a viabilidade do que foi exposto na rotina de um laboratório de pesquisa. A seleção de uma amostragem de dados que satisfaçam uma condição na busca por causas comuns a efeitos aleatórios em espaços de tempo determinados e consideravelmente grandes levam o pesquisador a um esforço desumano, sem o uso de uma forma eficiente de estocagem, recuperação e análise da informação. Para completar, considere que o desgaste físico e mental aumenta a propensão a erros.

A ciência da computação desenvolveu um conceito de sistema para aplacar tais necessidades: Sistemas de Banco de Dados. "Sistemas Gerenciadores de Bancos de Dados [ou SGBD] são conjuntos de dados associados a conjuntos de programas para acesso a esses dados. O conjunto dos dados é comumente chamado de banco de dados" (SILBERSCHATZ e KORTH, 1999:1).

Armazenar dados em SGBDs ao invés de armazená-los em um arquivo texto ou planilhas eletrônicas traz uma série de vantagens. Para que as informações possam entrar em um SGBD, as mesmas devem cumprir um pré-requisito. O pré-requisito é *ser compatível com uma estrutura de armazenamento que cumpre restrições* que forcem os dados a uma organização a representar, de maneira mais fiel, a realidade. A estrutura de armazenamento, (que deve ser previamente criada antes do armazenamento de dados) é projetada de acordo com ambiente que produz os dados e precisa ser validada em dois modelos: primeiramente em um modelo conceitual e, em seguida, em um modelo físico. Segundo COGO (1997:7), "modelo é a representação abstrata e simplificada de um sistema real, com a qual se pode explicar ou testar seu comportamento, em seu todo ou em partes". Uma vez que esta estrutura de armazenamento é validada, a mesma é transformada em um código e este, ao ser inserido dentro do SGBD, cria a estrutura definida pelo projetista para abrigar os dados em seu interior, sempre respeitando regras de validação declaradas nos dois modelos anteriores. Planilhas eletrônicas e arquivos textos não fornecem estes benefícios.

O desenvolvedor inicia a criação da estrutura de armazenamento inserindo objetos do mundo real em um modelo conceitual. O modelo conceitual, segundo COGO(1997:28), é aquele em que os objetos, suas características e relacionamentos equivalem a retratação fiel do ambiente do mundo real que se pretende representar. CHEN (1976) elaborou uma forma amplamente utilizada nos dias de hoje, de estabelecer o modelo conceitual chamado de modelo Entidade-Relacionamento. O modelo Entidade Relacionamento agrupa todos os objetos que compartilham características comuns, chamadas atributos, em conjuntos distintos chamados entidades. Este modelo aplica as primeiras normas de validação à estrutura de armazenamento, as chamadas restrições de integridade. Segundo SILBERCHATZ E KORTH (1999:28) o modelo conceitual de CHEN (1976) leva a dois dos mais importantes tipos de restrições de integridade, a existência de dependências, onde se verifica se dados de um conjunto dependem ou não de dados de outros conjuntos para poderem ser inseridos, e o mapeamento de cardinalidades, onde se verifica, caso identificada uma relação lógica entre elementos de conjuntos distintos, a quantos elementos de outra entidade um dado elemento da entidade pode relacionar-se. Outra restrição de integridade importante, não explicitamente citada por SILBERCHATZ e KORTH (1999) é a individualidade, capacidade de distinguir ou não elementos de um dado conjunto entre si através da análise de uma ou mais qualidades agrupadas. Esta qualidade ou qualidades agrupadas que distinguem um elemento dos outros dentro do conjunto recebem o nome de chaves candidatas.

O modelo físico, segundo modelo ao qual a estrutura de armazenamento proposta deve ser adequada, é criado a partir do projeto em modelo conceitual. Segundo COGO(1997:30) o modelo físico corresponde a adequação do modelo conceitual às limitações do modelo matemático (a teoria dos conjuntos) utilizado pelo SGBD para representar as informações armazenadas, e as adequações às limitações de hardware do sistema computacional. Aqui com certeza existe a obrigação de se impor mais dois tipos de restrições: as restrições de domínio e as restrições de implementação. COGO(1997:255) diz que restrições de domínio são aquelas que definem o conjunto de valores possíveis para cada atributo de cada elemento de um dado conjunto. SILBERCHATZ e KORTH (1999:62) ressaltam a importância desta restrição implicitamente e define a noção de conjunto, no modelo relacional como subconjunto do produto cartesiano de uma lista de domínios. Restrições de implementação são adaptações ao hardware e ao modelo de Sistema de Banco de Dados utilizado. Pode-se citar por exemplo 3 modelos de Sistemas de Bancos de Dados: Relacional, Pós-Relacional e Orientado a Objetos. Este trabalho utiliza um sistema de banco de dados baseado no primeiro. Grande parte dos SGBDs modernos são construídos com o projeto embasado no modelo relacional descrito no trabalho de Edgar Frank Codd (1923-2003), mais especificamente suas 12 regras, segundo MACHADO (1996),

- 1-Toda informação num banco de dados relacional é apresentada a nível lógico por valores tabelas;
- 2- Todo dado em um banco de dados relacional tem a garantia de ser logicamente acessível, recorrendo-se a uma combinação do nome da tabela, valor de chave e o nome da coluna;
- 3- tratamento sistemático de valores nulos(ausência de dados);
- 4- O dicionário de dados(catálogo) relacional ativo é baseado no modelo relacional;

- 5- O SGBD relacional deve ter uma linguagem para definição, detalhamento e manipulação de dados;
- 6- Tratamento das atualizações de visões dos dados;
- 7- tratamento de alto nível para inserção atualização e eliminação dos dados;
- 8- Independência dos dados físicos (mudança na memória e métodos de acessos);
- 9- Independência de dados lógicos(mudança de qualquer tipo nas tabelas básicas, exe: divisão de uma tabela por linha e coluna);
- 10- Independência das restrições de integridade;
- 11- Independência de distribuição;
- 12- Não subversão das regras de integridade ou restrições quando se utiliza uma linguagem de baixo nível.

Segundo MACHADO(1996), muitos SGBDs modernos cumprem pelo menos 10 destas regras, sendo que os que as cumprem se enquadram em um modelo de sistemas de bancos de dados chamado relacional. Portanto um sistema gerenciador de banco de dados relacional garante, além de toda fidelidade de representação imposta aos dados pelas restrições presentes na estrutura de armazenamento no ato de estocagem (exigência da décima segunda regra de Cood), a relativa simplicidade de recuperação e atualização dos mesmos, exigidas pelas segunda e sétima regras de Cood. Por fim, a grande maioria dos SGBDs interagem com o mundo exterior através de uma linguagem padronizada, como solicitado pela quinta regra de Cood, o código de armazenamento a partir do modelo físico citado acima, a linguagem SQL ou Structured Query Language, criada pela IBM no início da década de 1970, especificada pelo comitê ISO (ANSI-ISO-9075-1-1999). Esta linguagem de consulta é, conforme a quinta regra de Cood, subdividida em dois grupos de subcomandos. O subconjunto SQL DDL, acrônimo para *data definition language* que corresponde os códigos para criar a estrutura de armazenamento dentro do SGBD e o subconjunto SQL-DML, acrônimo para *data manipulation language* que corresponde aos códigos para recuperar, alterar e remover informações gravadas na estrutura previamente criada.

O que motivou este trabalho foram as seguintes necessidades: proposta de padronização do processo de estocagem, contribuição e participação de toda comunidade científica no processo de padronização, possibilidade de criação de uma base de dados compartilhada, aproveitamento de esforços anteriores de objetivos comum aos deste trabalho; possibilidade de correção de séries de dados defeituosas por modelos de estimativa aplicando os vários modelos disponíveis criados pela comunidade científica; validação de hipóteses levantadas a partir de padrões de comportamento periódicos de origem climática terrestre ou extra-terrestre. Portanto, este trabalho propõe um sistema de nome Aton para automatizar o processo de aquisição e formatação de dados gerados no dataloger Campbell, possibilitar armazenamento local dos mesmos em um Sistema de Banco de Dados Relacional para análise de séries (e sua eventual correção através de modelos de estimativas) e oferecer possibilidade de concatenar bases de dados locais com bases de dados importadas através de cooperação com o sistema de compartilhamento de dados OpenDap, este por sua vez desenvolvido por participantes do programa de metas unidata pertencente ao consórcio UCAR.

2. METODOLOGIA

Em sua primeira versão, o sistema extrai os dados a partir de arquivos textos na forma ASCII com caracter de recuo de linha. Estes arquivos texto são gerados a partir de um dataloger Campbell 23 X e tem dados gravados em colunas dispostas segundo uma programação prévia. O programa Aton reconhece, mediante a criação de um perfil, qual sensor os dados de uma coluna representa em determinada data, bem como seu respectivo coeficiente de calibração na determinada data em que foi gravado o ultimo perfil em relação a data da medida lida. Os dados lidos são por sua vez são transformados em linguagem SQL-DDL compatível com o modelo estrela proposto por KIMBALL (1996), uma vez que o mesmo simula a existência de matrizes multidimensionais usadas no padrão netCDF em sistemas relacionais, que por sua vez são limitados a matrizes de dimensão 2, as tabelas. A estrutura estrela proposta para alocar os dados do dataloger deve ser compatível com a existente no servidor openDAP SQL Database Server and the Test Server versão 1.1.7, uma vez que este já se encarrega de fazer a tradução de dados em formato estrela de KIMBALL (1996) para o formato netCDF de maneira transparente.

Os dados de radiação solar global, direta e difusa foram fornecidos pelo laboratório de Radiometria Solar do Departamento de Recursos Naturais - Setor Ambientais – da Faculdade de Ciências Agrônomicas da UNESP/Botucatu. As coordenadas geográficas do local de medidas são: 22,85° graus de latitude, 48,45° de longitude e 786m de altitude. O período utilizado para estocagem de dados compreende os anos de 1996 a 2006, num total de 11 anos de medidas arquivadas em formato texto ASCII.

O sistema de Banco de Dados utilizado para interagir com esta versão do sistema Aton foi o MySQL, Sistema de Banco de Dados Relacional de código aberto com o projeto atualmente gerenciado pela Sun Microsystems. Para desenho do projeto do modelo conceitual e físico da estrutura de banco de dados do sistema Aton, foi utilizado o software Sybase PowerDesigner versão 11. O sistema foi desenvolvido utilizando a linguagem de programação C++ ISO/IEC 14882:1998 através do ambiente de desenvolvimento Microsoft Visual C++ versão 9. Para que o programa Aton pudesse acessar o SGBD MySQL, foi utilizada a biblioteca livre MySQL++ disponível para download em <http://tangentsoft.net/mysql++/>. Para criar um arquivo de parâmetros de funcionamento em formato XML, foi utilizada a biblioteca livre minixml, disponível para download em <http://www.minixml.org/>. Para poder utilizar as bibliotecas livres ao projeto Aton através do ambiente Microsoft Visual Studio 9, foi utilizado o software construtor de projetos livre kitware CMAKE versão 2.6 para Windows, disponível em <http://www.cmake.org/HTML/index.html>. Em uma fase

posterior de desenvolvimento, foi utilizado o ambiente de desenvolvimento Borland c++ Builder versão 2006. O sistema é composto por 4 módulos: Importação, estocagem, análise e qualidade. O fluxo de informação pode ser observado na figura 1:

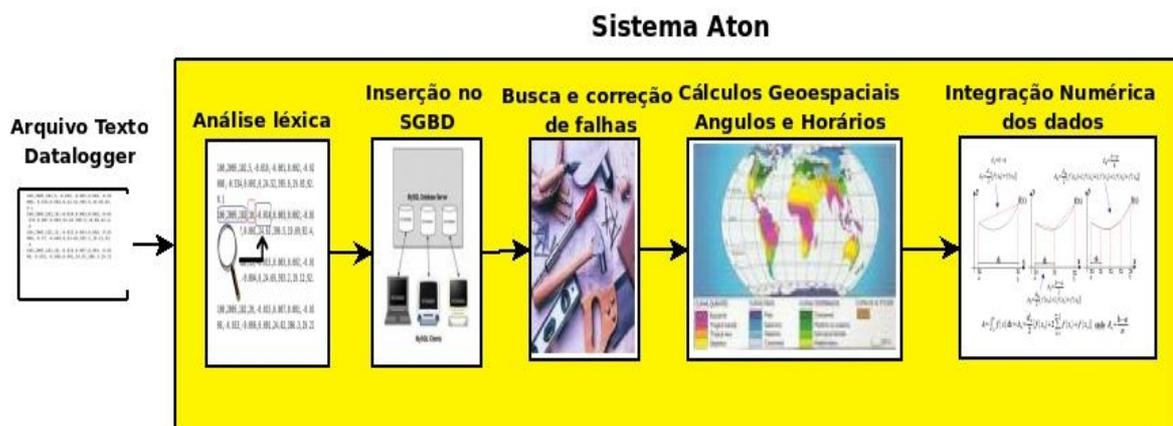


Figura 1 - Fluxo de informações no sistema Aton.

O módulo de importação localiza o diretório que contém todos os arquivos textos gerados pelos datalogger, nomeados de forma seqüencial e com extensão predefinida no arquivo de parâmetros e agrupa-os em um único arquivo temporário de análise. Este último será então lido e seu conteúdo de cada linha interpretado com base na data e horário da medida presente na linha. Uma vez identificada a data, hora e a posição de cada valor no arquivo de leitura temporário, todas as medidas obtidas a partir da linha de texto lida, bem como a ordem de suas posições nas colunas são armazenadas e enviadas ao módulo de estocagem, que é então solicitado.

O módulo de estocagem tem por objetivo deixar os dados de medidas de sensores compatíveis com o Sistema Gerenciador de Banco de Dados, interpretar a data, reconhecer o sensor da respectiva coluna e sua posição de onde a medida foi tomada no arquivo texto, e solicitar ao SGBD o valor da constante de calibração do sensor pertinente a data da medida, para que o valor de grandeza elétrica possa ser transformada em valor energético. Ao final de todo este processo, a medida pode então ser estocada no SGBD, o que é feito. Todo este processo é repetido para cada linha obtida do arquivo texto temporário.

O módulo de análise é solicitado quando se deseja recuperar e pesquisar informações do SGBD. Se constatado que a série de dados apresenta falhas, como lacunas ocasionadas por inoperabilidade temporária do sensor, então o usuário pode solicitar o módulo de qualidade.

O módulo de qualidade tem por objetivo sanar um dos grandes entraves apresentados no cotidiano de análise de dados relacionados a radiação solar é possível falha temporária de algum dos sensores de radiação e a conseqüente interrupção da aquisição de dados do respectivo sensor. Em modelos de estimativa, uma falha de amostragem pode comprometer a precisão quanto o comportamento da série de dados obtidos. Da mesma forma, uma integração, para uma análise energética em intervalos de tempo maiores só será possível se não existirem lacunas entre as somas no intervalo tomado. Assim sendo o módulo de qualidade, ao identificar séries de dados defeituosas, aplica um algoritmo que é semelhante ao método de interpolação entre pontos como enunciado abaixo:

“ Seja A ultimo valor energético válido antes do intervalo e B o próximo valor energético válido após o intervalo. Se o intervalo de tempo entre A e B é pequeno em relação ao intervalo de uma hora e a diferença do valor energético entre A e B for irrisória, então pode-se completar as lacunas de medida com o valor médio entre A e B”

“ Seja A ultimo valor energético válido antes do intervalo e B o próximo valor energético válido após o intervalo. Se o intervalo de tempo entre A e B é pequeno em relação ao intervalo de uma hora e a diferença do valor energético entre A e B for considerável, deverá ser utilizado um modelo de estimativas para preencher os valores”

A função de compartilhamento dos dados bem como sua concatenação é delegada aos softwares clientes OpenDAP do programa unidata.

3. RESULTADOS

Após uma importação feita pelo Sistema Aton de arquivos gerados pelo datalogger campbel da estação de radiometria solar de Botucatu, cujo formato pode ser verificado na figura 2, os resultados obtidos no interior do SGBD MySQL demonstrados na figura 3 puderam ser averiguados com a ajuda do software cliente MySQL Query Browser, software este independente do sistema Aton.

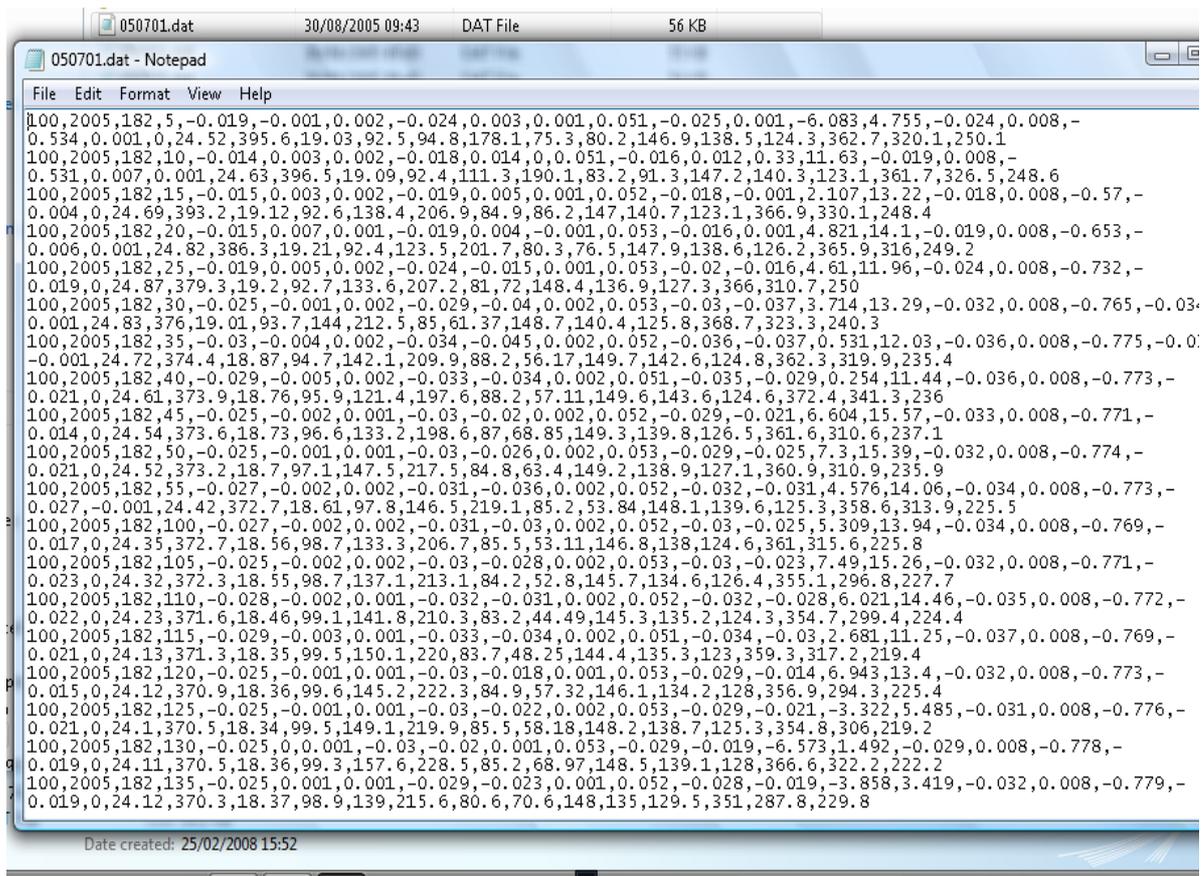


Figura 2 – Dados em formato texto gerados pelo Dataloger Campbell



Figura 3 – Base de dados de radiometria solar obtida pelo sistema Aton.

O módulo de importação condensou a medida de cada instante de cada sensor em uma linha na tabela `aton_facts_measures`, contendo o valor elétrico obtido do sensor e o valor energético obtido pela constante de calibração pertinente ao sensor cadastrada na última data em que foi feita a calibração em relação a data da medida processada.

Através do módulo de análise, o usuário constatou que no sensor de radiação direta, havia uma falha de leitura acusada pelo sistema Aton entre 10:20:00h e 10:35:00 horas graças ao sistema Aton, através de um parâmetro ajustável, reconhecer o código de erro -6999 programado no dataloger Campbell, como pode ser observado na figura 4.

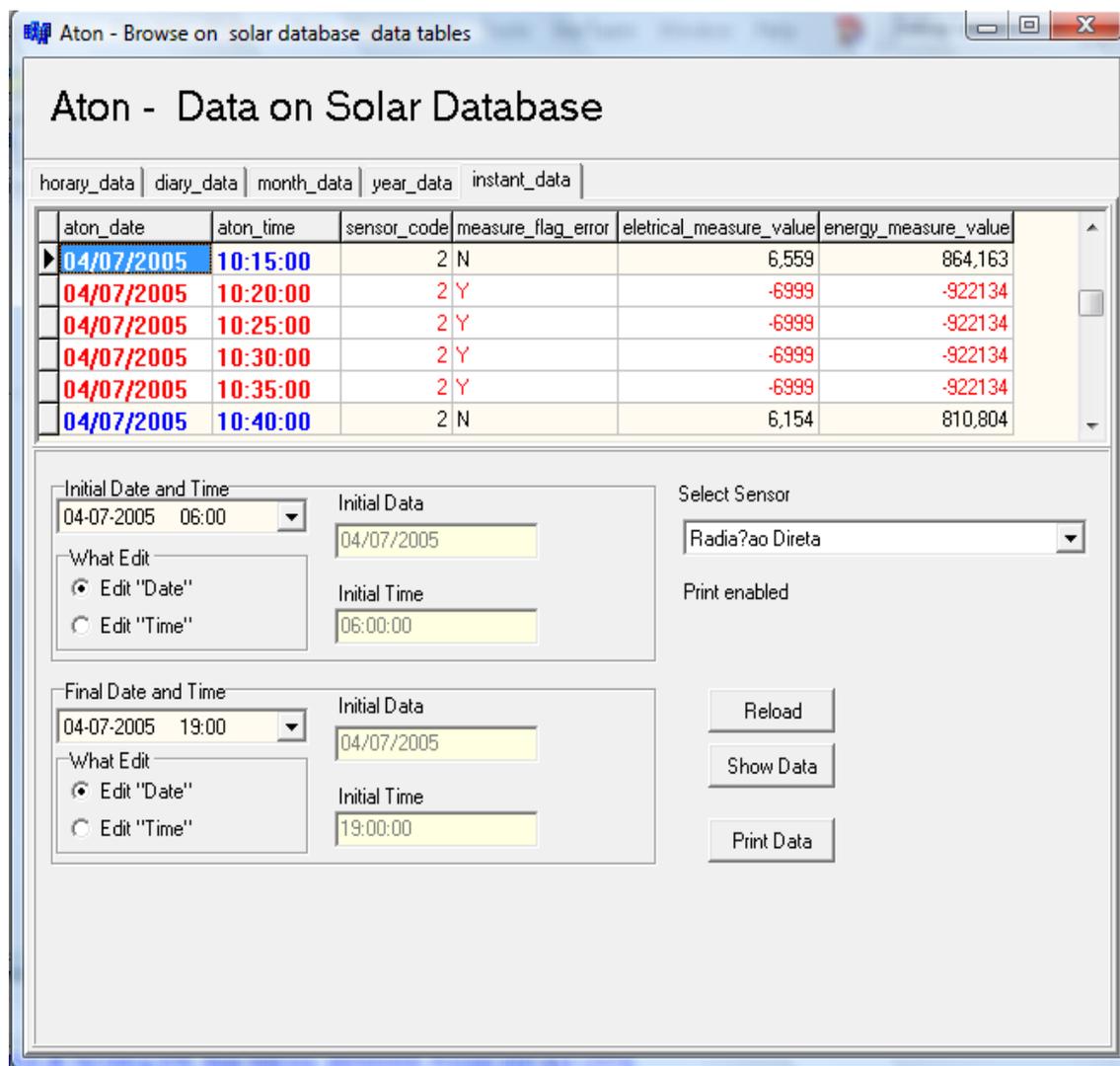


Figura 4 – Sistema Aton – Cliente Ra – Falha de leitura no sensor da radiação direta entre 10:20 e 10:35 do dia 4 de julho de 2005

4. CONCLUSÃO:

O sistema Aton, ao integrar dados não padronizados a um Sistema de Gerenciamento de Banco de Dados Relacional, e ao mesmo tempo, identificar sua confiabilidade e tornar-los compatíveis com softwares de compartilhamento de dados baseados em formatos científicos reconhecidos como o openCDF, abre uma possibilidade de integração de bases de dados de radiometria solar com outras bases de dados correlatas da climatologia através do sistema OpenDAP.

5. REFERÊNCIAS:

- Baltuch, M. S. 1997 Unidata's Internet Data Distribution (IDD) System: Two Years of Data Delivery. American Meteorological Society:1997 Disponível em: < <http://www.unidata.ucar.edu/software/idd/iips97.mitch.html> > acessado em: 20 abr. 08
- Chen, P. P. 1976 The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS)* New York, NY, USA, v. 1, n. 1, p. 9-36, 1976.
- Cogo, P. 1997 *Modelagem Conceitual e projeto de Banco de Dados*. 1 ed. Rio de Janeiro: Campus, 1997.
- Dal Pai, A., Escobedo, J. F. 2003 Equações de correção anisotrópica para a irradiância difusa medida por anel de sombreamento em função da cobertura de céu – I Modelagem. . In: XIII CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, Santa Maria – RS, Brasil. CD-ROM Anais XIII CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 3 a 7 de agosto de 2003.
- Date, C. J 2000. *Introdução a Sistemas de Bancos de Dados*. 7 ed. São Paulo: Campus, 2000 p. 21 – 82.

- Inmon, W. H. 2001 *Data Warehousing – Como transformar informações em oportunidades de negócios*. Tradução Melissa Kassner. 1 ed. São Paulo: Berkley, 2001
- Kimball, 1996 Ralph, *The Data Warehouse Toolkit*, John Wiley & Sons Inc., USA,
- Machado, F. N. R. 1996 *Projeto de Banco de Dados: uma visão prática*. 13 ed. São Paulo: Érica, 1996 p. 183 – 184.
- Rew,R; Davis, G 1990 NetCDF: an interface for scientific data access. *Computer Graphics and Applications, IEEE* v. 10 Issue 4 p. 76-82 Jul 1990
- Silberchatz, A. ; Korth, H. F. ; Sudarshan, S. 1999 *Sistema de Banco de Dados*. Tradução Marília Guimarães Pinheiro 3 ed. São Paulo: Makron Books, 1999
- Wikipédia, 2008- acessada em <http://pt.wikipedia.org/wiki/Esquema_do_banco_de_dados>

ATON- A SOLAR RADIOMETRY DATA AQUISITION, MANAGEMENT AND SHARING SOFTWARE.

Abstract. *This document presents Aton System development, a open-source software for solar radiometry aquisition, management and sharing data in compatible interface with cientific database integration software OpenDap. Aton have other features like maintenance of amount of damaged data by mecanical or eletrical causes in sensors or datalogger.*

Key words: Solar Energy, SQL , OpenDap,netCDF,Aton