

## ORGANIZAÇÃO E CONSISTÊNCIA DE BANCO DE DADOS PLUVIOMÉTRICOS DIÁRIOS

Jurandir Zullo Junior<sup>1</sup>, Hilton Silveira Pinto<sup>1</sup>, Débora Silveira Pinto<sup>1</sup>

**ABSTRACT** – This work has the objective to describe the organization and updating of a daily rainfall database considering its use in agroclimatic modeling. Daily data were sent by DAEE to Cepagri/Unicamp in different formats and times since 1987, having been processed in several computer systems. The data used to update the database in 2005 were obtained directly in the DAEE's webpage using a manual process. The database has now 1,528,393 records, 129.7MB and 49,303 years of daily data for 1,208 ground rainfall stations. The methods used to recover missing values and test daily data can be used to organize and update databases similar to this one.

### INTRODUÇÃO

A disponibilidade de séries históricas de dados pluviométricos diários consistidos, com qualidade e confiabilidade, é fundamental para a realização de trabalhos na área de modelagem agroclimatológica. Como exemplo, cita-se os trabalhos de zoneamento agrícola que são baseados na utilização de séries históricas, normalmente com duração mínima de 10 anos e desejável de pelo menos 30 anos. As perdas agrícolas no Brasil, até a primeira metade da década de 90, eram devidas à falta de chuvas na fase crítica das culturas (60% dos casos) e ao excesso na colheita (em 30% das situações) (Rossetti, 2001). Ou seja, elas estavam diretamente relacionadas, entre outros fatores, ao pouco conhecimento que os agricultores possuíam do regime hídrico em sua região.

O Brasil, devido à sua extensão territorial, à dificuldade de processamento e armazenamento de grandes volumes de dados, da ordem de terabytes, e ao alto custo relacionado à instalação e manutenção de estações meteorológicas, não é diferente de grande parte do globo no que se refere à disponibilidade de redes meteorológicas de superfície que sejam consistentes, confiáveis e adequadas para utilização na modelagem agroclimatológica. A popularização da Internet, a melhoria dos sistemas de gerenciamento de bancos de dados e do processamento de alto desempenho e o aumento da disponibilidade de sistemas de computação robustos e de baixo custo estão contribuindo substancialmente no que se refere ao processamento e armazenamento dos dados meteorológicos. Como exemplo, cita-se o Agritempo (Evangelista et al., 2003) que é um sistema de monitoramento agrometeorológico e agroclimatológico baseado em tecnologia moderna de computação e que tem, entre outras finalidades, armazenar e colocar à disposição, da melhor forma possível, as séries históricas de dados climáticos. Isto, sem dúvida, é um ganho ao que se tinha até a alguns anos atrás quando grande parte dos dados meteorológicos e climáticos praticamente não estava disponível em mídia eletrônica.

Além disso, as instituições governamentais e privadas têm empreendido significativos recursos nos últimos anos visando o aumento do número de estações meteorológicas de superfície e substituição

dos modelos mecânicos pelos automáticos. Contribuem para este esforço as tentativas de redução de custos das estações meteorológicas, especialmente as automáticas. O aumento do número de estações de superfície produz um impacto direto, em curto prazo, nos sistemas de monitoramento agrometeorológico, pois permitem que eles sejam mais precisos e detalhados. No caso de estudos climáticos, o impacto do aumento da quantidade de estações de superfície é em médio prazo, pois é preciso que haja um acúmulo de pelo menos 10 anos de dados para que ela seja útil.

Dentro deste contexto, é importante ressaltar o trabalho realizado pelo Departamento de Águas e Energia Elétrica (DAEE), órgão gestor dos recursos hídricos do Estado de São Paulo, que opera uma rede de estações pluviométricas com mais de 1.000 postos distribuídos pelo Estado. Pelo menos um terço destas estações possui, no mínimo, 30 anos de dados diários sem falhas. Trata-se, portanto, de um banco de dados de grande utilidade para a realização de estudos climáticos no Estado de São Paulo, desde que devidamente consistido e testado. Embora boa parte deste banco esteja disponível atualmente na página internet do DAEE, é importante ressaltar que grande parte dos dados é apresentada como não tendo sido consistida. Por essa razão, este trabalho apresenta o método utilizado no Cepagri/Unicamp para a consistência dos dados pluviométricos do DAEE/CTH, visando a utilização de séries históricas em trabalhos de modelagem agroclimatológica.

### MATERIAL E MÉTODOS

Os dados pluviométricos utilizados neste trabalho foram obtidos no banco de dados disponível no Cepagri/Unicamp e na página Internet do DAEE (em <http://www.daee.sp.gov.br>). O banco de dados existente no Cepagri/Unicamp começou a ser implantado em 1987, a partir da cessão de uma série de dados gravados em fitas magnéticas de rolo. O processamento naquela época era realizado em computadores do tipo "mainframe" com acesso via terminal remoto e apresentação dos resultados, predominantemente, em formulários de papel. O grande tamanho do banco para a época, algo em torno de 150MB, requereu que ele fosse implantado e processado em vários sistemas, de acordo com a necessidade e disponibilidade de computadores na Unicamp. Entre eles, cita-se o VAX 780/Digital, o CBERS/CDC, o IBM 3090 e a Workstation Sparc Sun. A partir da década de 90, os dados de atualização do banco começaram a ser enviados em disquetes, e não mais em fitas magnéticas de rolo, como anteriormente. Na segunda metade da década de 90, com a evolução tecnológica dos microcomputadores pessoais, foi possível fazer a migração do banco para este tipo de sistema. Isto facilitou o processamento e atualização do banco, a disseminação dos dados disponíveis e a utilização de outros sistemas, especialmente as planilhas eletrônicas e os gerenciadores de bancos de dados.

<sup>1</sup> Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI), Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, 13083970, Campinas, SP ([jurandir@cpa.unicamp.br](mailto:jurandir@cpa.unicamp.br), [hilton@cpa.unicamp.br](mailto:hilton@cpa.unicamp.br), [debora\\_cpa@cpa.unicamp.br](mailto:debora_cpa@cpa.unicamp.br)).

O banco original, utilizado neste trabalho, contém dados a partir de 1934, para 1.339 postos, totalizando 127,8MB, 1.505.794 registros e 48.574 anos. A montagem deste banco foi feita em várias etapas. Na primeira delas, realizada em 1988, foram enviados seis arquivos com dados entre os anos de 1934 e 1987 para 1.635 postos, totalizando 130,4MB, 44.096 anos e 1.366.976 registros. Neste total já não estão contabilizados 6.856 anos que foram enviados em duplicidade, correspondendo a 252 postos.

Na segunda etapa, ainda em 1988, foi enviado mais um arquivo com dados entre 1986 e 1988 para 1.011 postos, totalizando 3.028 anos e 8,9MB. Foram enviados 1.006 anos em duplicidade. O arquivo-base, nesta época, possuía dados para 1.644 postos, totalizando 132,1MB, 1.399.681 registros e 45.151 anos. Este arquivo passou por um primeiro processo de consistência quando foram eliminados os dados de 340 postos, correspondendo a 3.832 anos, que tinham as seguintes características:

- a) série menor que 10 anos de dados consecutivos e sem falhas, terminando antes de 1985, inclusive;
- b) localização desconhecida.

Em seguida, foi realizado um segundo processo de consistência, quando foram eliminados 4.921 anos que tinham as seguintes características:

- a) pelo menos dois meses sem dados;
- b) pelo menos dois meses da estação chuvosa (outubro a março) sem registro de chuvas;
- c) pelo menos um mês sem dados e um mês da estação chuvosa sem registro de chuvas.

Estes testes foram feitos assumindo que até a falta de um mês de dados poderia ser devida ao período de férias do operador do posto. Entretanto, mais que isso poderia indicar problemas sérios com o posto que colocaria em dúvida a confiabilidade e a consistência de sua série histórica. A existência de mais de um mês do período chuvoso (outubro a março) com chuva nula também poderia indicar problemas no posto, pois a probabilidade deste evento ocorrer em São Paulo é muito baixa. Isto poderia ser devida à substituição do valor representativo da falta de dados (999.9) pelo valor nulo, na digitação ou no preenchimento das planilhas de dados. O arquivo-base resultante destas duas consistências totalizou 106,5MB para 36.398 anos, 1.128.338 registros e 1.304 postos.

Este arquivo passou, então, por um processo de substituição de dados diários faltantes, representados, originariamente, pelo valor 999.9. Os dados ausentes foram substituídos, se possível, pela média aritmética dos dados dos postos vizinhos. Por posto vizinho entende-se o posto localizado dentro de um círculo de raio igual a 0,0625 graus (aproximadamente 6,25km) com centro no posto cujo dado pretendia-se recuperar e que estivesse na mesma classe de altitude deste. Foram definidas as três seguintes classes de altitude: a) litoral (altitude menor ou igual a 300m), b) interior (altitude entre 300m e 850m) e c) serra (altitude superior a 850m). Com estas duas definições (distância máxima e classe de altitude) procurou-se evitar a utilização de dados obtidos em situações de altitude e localização muito distintas do posto cujo dado pretendia-se recuperar. Foram substituídos 77.750 dias faltantes, correspondendo a 0,6% do total de dados existentes no banco.

Após a formação deste arquivo-base, foram enviados pelo DAEE, em momentos distintos, mais sete arquivos de atualização correspondentes ao período de

1987 a 2000 que, seguindo os mesmos procedimentos de consistência e recuperação descritos anteriormente, resultaram em um novo arquivo totalizando 127,8MB, 1.505.794 registros para 1.339 postos. Em 1998, o DAEE publicou os dados do seu banco pluviométrico através de CD-ROM com atualização até 1997. Em seguida, colocou-os à disposição na sua própria página internet. Estes dados foram consultados e transferidos manualmente para o Cepagri/Unicamp totalizando 821 arquivos com dados até 2004 para alguns postos. Estes arquivos foram adequadamente combinados e formaram um novo consolidado contendo todos os dados pluviométricos diários disponíveis na página internet do DAEE, totalizando 38.145 anos, 100,4MB, 105.242.055 registros para 821 postos. Destaca-se que boa parte destes dados é assinalada no arquivo, pelo próprio DAEE, como não tendo sido consistidos. Isto ressalta a importância das consistências realizadas no arquivo-base à medida que foi sendo atualizado. Este arquivo, com dados até 2004, foi combinado com o arquivo-base anterior (possuindo dados até 2000) e passou pelos processos de consistência e recuperação descritas anteriormente gerando o arquivo-base atual.

## RESULTADOS E DISCUSSÃO

O arquivo-base, com dados até 2004, resultante dos processamentos descritos acima, contém 49.303 anos, 1.528.393 registros, 129,7MB para 1.208 postos. Trata-se de um banco de dados de grande utilidade para estudos agroclimáticos, especialmente considerando a existência de dados até 2004. A série mais extensa é de um posto localizado no município de Campinas e contém 62 anos de dados sem falhas, entre 1942 e 2004. A série mais antiga contém dados desde 1889, porém com anos faltantes. Ressalta-se que a opção sempre foi pela qualidade dos dados do banco ao invés da quantidade de dados no banco. Sendo assim, preferiu-se eliminar dados duvidosos a mantê-los armazenados. Certamente o número elevado de estações disponíveis também permitiu que se fosse mais rigoroso nas consistências realizadas. Além disso, observou-se, através de inspeção visual das séries diárias e mensais de vários postos, que, normalmente, eles eram muito bons (ou seja, tinham séries longas e com poucas falhas) ou muito ruins (tinham séries com menos de 10 anos e/ou com muitas falhas).

## REFERÊNCIAS

- Evangelista, S.R.M., Ternes, S., Assad, E.D., Romani, L.A.S., Santos, E.H.dos, Franzoni, A. Monitoramento Agroclimatológico pela WEB. IV Congresso Brasileiro da Sociedade Brasileira de Informática Aplicada à Agropecuária e à Agroindústria. Anais .. Porto Seguro, 2003.
- Rossetti, L.A. Zoneamento agrícola em aplicações de crédito e securidade rural no Brasil: aspectos atuariais e de política agrícola. Revista Brasileira de Agrometeorologia, Passo Fundo, v.9, n.3, p.386-399, 2001.